Leaf endophytic bacteria in *Plantago lanceolata* and the effects of soil and maternal genotype on the endophyte community

Joose Lassila

Master´s thesis

Faculty of Biological and Environmental Sciences

Master's Programme in Integrative Plant Sciences

University of Helsinki

2020

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty Faculty of Biological and Environmental Sciences		Koulutusohjelma – Utbildningsprogram – Degree Programme Master's Programme in Integrative Plant Sciences
Tekijä – Författare – Author Joose Lassila		
Työn nimi – Arbetets titel – Title Leaf endophytic bacteria in *Plantago lanceolata* and the effects of soil and maternal genotype on the endophyte community		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Plant Sciences		
Työn laji – Arbetets art – Level Master's Thesis	Aika – Datum – Month and year 4/2020	Sivumäärä – Sidoantal – Number of pages 53

Tiivistelmä – Referat – Abstract

The aim of this study was to examine the leaf endophytic bacteria in *Plantago lanceolata*. The first aim was to get a comprehensive picture of the bacterial diversity, by screening for the different bacterial genera inside the leaves. Furthermore, I aimed to examine the effect of soil and maternal genotype on the endophytic community within *P. lanceolata* leaves and search for clues of vertical inheritance of endophytes from parent to offspring via seeds. I studied the endophytic bacteria by extracting DNA from the plant leaves and by trying to amplify any bacterial DNA present to get a view of the bacterial diversity in the leaves. My aim was to compare the bacterial community of the mother plants to that of their offspring and also to compare the bacterial communities of plants grown in different soil conditions. Furthermore, I tried to study how the soil conditions affect the growth of *P. lanceolata* seedlings.

I collected seeds and leaf samples of *P. lanceolata* from Åland, Southwestern Finland, from a population that is part of the ongoing long-term metapopulation research started in Åland in the early 90's. I marked 21 plant individuals (hereafter referred to as the "mother plants") in the field in June, when collecting the first leaf samples. In August I collected all seeds from the same plant individuals and a second set of leaf samples. I also collected soil samples from the same location. With the seeds collected from the wild population I executed a growth experiment in Viikki, Helsinki. I grew one set of seeds in twice autoclaved sand (hereafter referred to as the "sterile soil") and another set in twice autoclaved sand mixed with soil collected from the Åland population (hereafter referred to as the "environmental soil"). I surface sterilized all seeds and then sowed each in their own growth pot and placed them in a growth chamber. During the experiment I took measurements of the leaves. At end of the growth experiment, I took samples of the leaves and surface sterilized them to exclude any epiphytic microorganisms from the analysis. I also surface sterilized the leaf samples taken from the mother plants. I then extracted DNA from the leaf samples and run PCR to amplify certain regions of the bacterial 16S rDNA gene, that is widely used in bacterial taxonomy. The obtained DNA reads where then clustered into Operational Taxonomic Units (OTUs) and assigned taxonomy using SILVA reference database.

Mitochondria and chloroplasts of eukaryotic organisms also harbour 16S rDNA regions, so the challenge of studies looking at endophytic bacteria is to exclude the 16S regions of mitochondria and chloroplasts. This proved to be a problem in my study also. More than 86% of all DNA reads obtained turned out to be from *P. lanceolata* mitochondria and more than 12% from *P. lanceolata* chloroplasts. Only a bit more than 1% of the reads were eubacterial. This effectively hindered reliable analysis of the endophyte community. I nevertheless analysed the observed eubacterial diversity although the results must be taken as only preliminary and with utmost caution. The eubacterial reads clustered into 218 OTUs, representing 71 different bacterial genera. Six most common genera constituted over 83% of eubacterial reads. Most of these bacteria, most notably *Shewanella*, *Ralstonia* and *Halomonas*, could be identified as being clearly contaminants and not real endophytes. For the 65 less common bacterial genera I performed community analysis using Bray-Curtis Dissimilarity index and Analysis of Similarities (ANOSIM). The results showed that there was a significant difference between the different soil treatments ($P = 0.014$, $R = 0.3787$) and also between the two growth chambers ($P = 0.011$, $R = 0.5493$). I found no effect of maternal genotype on the bacterial community. Therefore, I observed no sign of vertical inheritance of endophytes. The growth experiment results showed that germination percentage was significantly lower in the environmental soil than in the sterile soil for all genotypes ($F = 10.78$, $P = 0.0012$). However, seedling in the environmental soil grew bigger than the seedlings in the sterile soil ($F = 10.91$, $P < 0.0001$).

For future studies on similar topics, validating molecular methods before large scale sequencing could yield more reliable results. Size fractionating the DNA products of the first PCR round could exclude most mitochondrial sequences and hence allow better analysis of endophytes. This would enable studying interesting questions on coevolution and ecology of host-endophyte interactions. Although I did find some differences in the bacterial communities of different treatments, these results must be considered with caution and as only preliminary.

Avainsanat – Nyckelord – Keywords
Plantago lanceolata; Åland; microbiome; bacteria; endophytes; plants; coevolution

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Anna-Liisa Laine, Riitta Nissinen, Hanna Susi

Säilytyspaikka – Förvaringsställe – Where deposited
HELDA - Helsingin yliopiston digitaalinen arkisto / HELDA - Helsingfors universitets digitala publikationsarkiv /HELDA - Digital Repository of the University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty Bio- ja ympäristötieteellinen tiedekunta	Koulutusohjelma – Utbildningsprogram – Degree Programme Kasvitieteen maisteriohjelma	
Tekijä – Författare – Author Joose Lassila		
Työn nimi – Arbetets titel – Title Heinäratamon lehtiendofyyttiset bakteerit ja kasvualustan ja genotyypin vaikutus endofyyttiyhteisöön		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Kasvitiede		
Työn laji – Arbetets art – Level Maisterintutkielma	Aika – Datum – Month and year 4/2020	Sivumäärä – Sidoantal – Number of pages 53

Tiivistelmä – Referat – Abstract

Tässä tutkielmassa tarkastelen heinäratamon (*Plantago lanceolata*) lehtien endofyyttibakteereita. Työn ensimmäinen tavoite oli saada yleiskäsitys heinäratamon lehtiendofyyttien monimuotoisuudesta kartoittamalla eri bakteerisukujen esiintyminen heinäratamossa. Lisäksi pyrin selvittämään bakteeriyhteisöön vaikuttavia tekijöitä, erityisesti kasvin genotyypin sekä kasvualustan vaikutusta. Tarkoitukseni oli selvittää, onko löydettävissä viitteitä endofyyttibakteerien vertikaalisesta periytymisestä sukupolvelta toiselle siementen mukana ja vaikuttaako kasvualusta, erityisesti kasvualusta luonnollinen mikrobisto, endofyyttiyhteisöön. Selvitin asiaa eristämällä kasvien lehdistä DNA:ta ja monistamalla siitä bakteeri-DNA:n jaksoja. Tarkoitukseni oli verrata äitikasvien lehtien endofyyttejä siemenistä itäneiden jälkeläisten lehtien endofyytteihin sekä verrata erilaisessa maakäsittelyssä kasvatettujen kasvien endofyyttiyhteisöjä. Lisäksi pyrin selvittämään, miten kasvualusta vaikuttaa heinäratamon siementen itämiseen ja kasvien kasvuun.

Keräsin heinäratamon siemeniä Ahvenanmaalta populaatiosta, joka on osa pitkäaikasta 90-luvulla aloitettua metapopulaatiotutkimusverkostoa. Alkukesästä merkitsin maaston 21 kasviyksilöä (äitikasvit) ja keräsin niistä lehtinäytteitä. Loppukesästä keräsin samoista kasviyksilöistä uudet lehtinäytteet sekä siemeniä. Samalla keräsin populaation ympäristöstä myös maa-ainesta. Siemenillä suoritin Helsingissä Viikissä kasvatuskokeen. Kasvatin osan siemenistä kahdesti autoklavoidussa hiekassa (”steriili maa”) ja toisen osan kahdesti autoklavoidussa hiekassa, johon sekoitin luonnosta keräämääni maa-ainesta (”luonnollinen maa”). Ennen kasvatuksen alkua pintasteriloin kaikki siemenet ja sitten kylvin jokaisen niistä omaan kasvatuspurkkiinsa ja asetin purkit kasvatuskaappiin. Kasvatuksen aikana tarkkailin kasvien itämistä ja mittasin lehtien kasvua. Kasvatuksen lopussa otin kaikista kasveista lehtinäytteet ja pintasteriloin ne poistaakseni mahdolliset epifyyttiset mikro-organismit. Samoin pintasteriloin äitikasveista otetut lehtinäytteet. Seuraavaksi eristin näytteistä DNA:ta ja monistin PCR:llä DNA:sta tiettyjä alueita prokaryoottisesta 16S rDNA -geenistä, jota käytetään laajasti bakteerien taksonomisessa ryhmittelyssä. Eristetyt sekvenssipätkät ryhmittelin operatiivisiin taksonomisiin yksiköihin (OTU) ja vertasin niitä SILVA tietokantaan taksonomian selvittämiseksi.

Koska myös eukaryoottisten eliöiden mitokondrioissa ja kloroplasteissa on 16S rDNA -alueita, niiden geenit voivat vaikeuttaa endofyyttibakteerien tutkimista. Tämä muodostui ongelmaksi myös tässä tutkielmassa. Yli 86 % kaikista monistamistani sekvenssipätkistä oli peräisin heinäratamon mitondrioista ja yli 12 % heinäratamon kloroplasteista. Vain hieman yli yksi prosentti oli bakteerialkuperää. Tämä käytännössä katsoen esti endofyyttiyhteisön luotettavan tarkastelun. Analysoin siitä huolimatta havaitut bakteeriyhteisöt, joskin tuloksia voidaan pitää lähinnä suuntaa antavina ja niihin on syytä suhtautua äärimmäisellä varovaisuudella. Bakteerisekvenssit jakautuivat 218 taksonomiseen yksikköön, jotka edustivat 71 erilaista bakteerisukua. Kuusi yleisintä sukua muodosti yli 83 % kaikista bakteerisekvensseistä. Useimmat näistä kuudesta suvusta, erityisesti *Shewanella*-, *Ralstonia*- ja *Halomonas*-suvut olivat selkeästi kontaminantteja eivätkä aitoja endofyyttejä. Tein yhteisöanalyysin 65:llä vähemmän yleisellä bakteerisuvulla. Käytin analyysissä Bray-Curtis -dissimilariteetti-indeksiä ja Analysis of Similarities (ANOSIM) -testiä. Havaitsin tilastollisesti merkitsevän eron eri maakäsittelyjen välillä ($P = 0.014$, $R = 0.3787$) sekä eri kasvatuskaappien välillä ($P = 0.011$, $R = 0.5493$). En löytänyt merkkejä genotyypin vaikutuksesta bakteeriyhteisöön. Siten en havainnut merkkejä myöskään endofyyttien vertikaalisesta periytymisestä äitikasveilta jälkeläisille. Kasvatuskokeen tulokset osoittivat, että itämisprosetti oli kaikilla genotyypeillä merkitsevästi matalampi luonnollisessa maassa kasvatetuilla siemenillä kuin steriilissä maassa kasvatetuilla ($F = 10.78$, $P = 0.0012$). Toisaalta luonnollisessa maassa kasvit kasvoivat suuremmiksi kuin steriilissä maassa ($F = 10.91$, $P < 0.0001$).

Analyysimenetelmien validointi olisi äärimmäisen tärkeää mahdollisissa jatkotutkimuksissa tästä aiheesta. Ennen suuren mittakaavan sekvensointia olisi syytä varmistaa, että mitokondrio- ja klorplastigeenit saadaan tehokkaasti poistettua analyyseistä. Tämän pitäisi onnistua PCR-tuotteiden kokofraktioinnilla, jolloin saattaisi olla mahdollista saada luotettavampi kuva bakteeriyhteisöstä. Tämä avaisi kiehtovia mahdollisuuksia tutkia endofyyttien ja isäntäkasvien koevoluutiota ja ympäristötekijöiden vaikutusta endofyytteihin ja toisin päin. Vaikka havaitsin joitain eroja eri käsittelyjen bakteeriyhteisöissä, on tämän tutkielman tuloksiin suhtauduttava hyvin varovasti ja korkeintaan suuntaa-antavina.

Avainsanat – Nyckelord – Keywords
Plantago lanceolata; Ahvenanmaa; mikrobiomi; bakteerit; endofyytit; kasvit; koevoluutio

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Anna-Liisa Laine, Riitta Nissinen, Hanna Susi

Säilytyspaikka – Förvaringsställe – Where deposited
HELDA - Helsingin yliopiston digitaalinen arkisto / HELDA - Helsingfors universitets digitala publikationsarkiv /HELDA - Digital Repository of the University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information

Contents

1 Introduction

1.1 Plants and microbes

Plants, like any other organisms, live in constant interaction with a myriad of microbes, including bacteria, viruses and fungi (Turner, James, Poole 2013). Microbes are omnipresent in the environment and hence, plants confront them during all phases of their life cycle and with all their organs: Soil microbes as a germinating seed and with roots, airborne and animal transmitted microbes with their aerial parts (Hardoim et al. 2015; Turner, James, Poole 2013). The interaction between an individual plant and microbes can be anything from mutualism to parasitism. Microbes that confer some sort of growth or health benefit for the plant, e.g in the form of extra nutrients, are generally regarded as mutualists whereas lethal disease-causing microbes are considered antagonists. The interaction can, however, also be anything between these two extremes. Since plants are in contact with an endless number of microbes and only some of them can be regarded as clearly mutualistic or antagonistic, it is likely that often the interaction between a plant and a microbe is somewhat neutral, without any evident influence on their fitness. Sometimes the outcome and magnitude of an interaction is also context-dependent and can change depending on environmental factors or other species present (Laitinen, Hellström, Wäli 2016; Marquez et al. 2007; Rubin, van Groenigen, Hungate 2017).

1.2 Soil microbes

Soil is the most important area where plants confront microbial organisms (Berendsen, Pieterse, Bakker 2012). The most biologically active part of the soil is the thin area around plant roots, termed the rhizosphere, in which plant root exudates enhance microbial life, a phenomenon termed the 'rhizosphere effect' (Raaijmakers et al. 2009). Plants are capable of attracting and choosing beneficial microbes in the rhizosphere, some of which establish stable symbiosis with plants, or even enter root tissues as endophytes. Furthermore, soil is an important source of plant pathogenic microbes: Rhizosphere is where free-living pathogens compete with other microbes and try to gain entry into plant tissues (Raaijmakers et al. 2009).

The importance of mutualistic mycorrhizal symbioses between plants and fungi for plant nutrition and health is well established and understood (Heijden et al. 2015). It is thought that fungi may have played a crucial role even in the life of the very first land plants (Gehrig, Schüssler, Kluge 1996; Pirozynski and Malloch 1975; Selosse, M-A and Le Tacon 1998). Mutualistic symbioses

between some plants, most importantly legumes, and nitrogen-fixing bacteria are also well documented (Franche, Lindström, Elmerich 2009; Masson-Boivin et al. 2009). These bacteria originate in the surrounding soil and they usually induce the formation of special root nodules in which they live as colonies. Nitrogen-fixing bacteria can also be found free-living in the rhizosphere or as endophytes. The bacteria fix atmospheric nitrogen, supplying the plant with extra nutrients and therefore enhance its growth. In general, symbiotic bacteria that enhance plant growth are referred to as Plant Growth Promoting Bacteria (PGPB) (Bashan and Holguin 1998). Another commonly observed mechanism of growth promotion is production of plant hormones or hormone-like substrates by bacteria or other microbes, which can result in enhanced growth and development, promotion of mutualistic symbiosis, e.g. mycorrhizal connections, and protection from biotic or abiotic stress (Ortíz-Castro et al. 2009). Some microbes can promote plant life by suppressing pathogens by direct inhibition or indirectly by competition or by priming plant immunity (Bhattacharyya and Jha 2012; Compant, Clément, Sessitsch 2010; Pieterse et al. 2014; Raaijmakers et al. 2009).

1.3 Hologenome concept and evolution

The focus of evolutionary studies is shifting more and more from an individual species to the "metaorganism" i.e. the focal species and all its symbionts. Metaorganism can be defined as one co-evolving unit of selection (Berg et al. 2016). From this perspective evolutionary selection works on the set of genomes, the hologenome, of all these closely interacting species (Rosenberg et al. 2010; Rosenberg and Zilber-Rosenberg 2016; Sánchez-Cañizares et al. 2017; Zilber-Rosenberg and Rosenberg 2008). In many cases this approach is well justified since we know that often symbiosis is crucial, even a prerequisite, for the survival of the organisms. The significance of gut microbiota for human immune system and health is well established (Cho and Blaser 2012) and root microbes have been regarded as plants' version of gut microbiota (Selosse, Marc-André, Bessis, Pozo 2014) or the "second genome of plants" (Berendsen, Pieterse, Bakker 2012). In this respect, there seems to be little sense in considering plants, or any multicellular organism for that matter, apart from their microbial symbionts. Close interaction between the host and its symbionts, if it has fitness consequences for the organisms, can result in coevolution between them. Coevolution can refer to reciprocal adaptation to the selection pressure caused by the presence of other organisms, as in a classic predator-prey interaction where both partners inflict opposing selection pressure on each other, which results in antagonistic coevolution between them. Coevolution in the case of tight mutualistic symbioses, can on the other hand mean, that organisms rely on each other on crucial metabolism or other functions. Occasionally such tight relationships have evolved that

neither of the partners can anymore live without the other. Genomic erosion or adaptive gene loss refers to a situation where (usually microbial) symbionts have lost genes responsible for some vital metabolic functions and rely solely on their host for providing the needed metabolites (Moran, McLaughlin, Sorek 2009; Morris, Lenski, Zinser 2012). This is particularly common among endophytic symbionts, but has also been reported from epiphytic microbes which then rely on their host community to provide the necessary gene products (Spanu et al. 2010).This way a community may in some aspects function as a "superorganism" with high degree of specialisation between its members. There are already well-established cases where the symbiosis is defined as a distinct organism, as in the case of lichens (Honegger 1991).

The origin of eukaryotic cells by endosymbiosis can be viewed as a prime example of symbiosis formation opening up new perspectives for evolutionary development (Sachs, Skophammer, Regus 2011). Symbiosis may lead to emergence of novel functions where the resulting combination of organisms working in synchrony (or "metaorganism") is more than the sum of its parts. When early eukaryotic cells acquired the endosymbionts that became mitochondria and chloroplasts, this paved way for the emergence of the vast diversity of eukaryotes and multicellular organisms that we know today. Furthermore, these remarkable evolutionary innovations have a great impact on ecosystems and status of the whole planet: production of oxygen as a side product of photosynthesis and sequestration of carbon has had huge impact on the composition of the atmosphere and hence to the basis of nearly all forms of life. To a lesser extent, symbiosis of plants and e.g. nitrogen fixing bacteria affects nutrient cycling and hence ecosystem functioning. The coevolution between insects and flowering plants has resulted in the adaptive radiation of both groups (Lunau 2004).

This view differs from the traditional view of evolution as a slow process of distinct individual species slowly evolving as a result of natural selection and mutation to their own genome. Rather, evolutionary change sometimes accelerates when formation of novel symbiotic relationships results in new functioning and adaptations of the metaorganism, enabling e.g. use of new energy sources or better adaptation to stress. Furthermore, evolutionary change sometimes gets new raw material not only from mutations, but also from horizontal gene transfer between not closely related species. Endosymbiosis can also result in permanent modification of the host and/or symbiont genome. Viruses often implement their genetic material into that of their host and this may sometimes result in the viral genes becoming part of the host genome, which is transmitted to the new generation: it has been estimated that even 10% of the genetic material of some vertebrate species is of viral origin (Stoye 2012). Symbionts on the other hand may experience genomic erosion or adaptive gene loss whereby they lose certain genes in a situation where

required gene products can be easily obtained from the host or the surrounding environment (Moran, McLaughlin, Sorek 2009; Morris, Lenski, Zinser 2012; Pinto-Carbó et al. 2016).

Since in nature organisms live in a complex network of interactions where the strength and significance of the interaction is a continuum rather that well-defined phenomenon, it is impossible to draw any clear border between a "close" and "distant" symbiosis. Hence the "metaorganism" can be a rather vague concept, and it may not be possible to clearly define it or to quantify its significance for evolution. However, this concept helps to appreciate the role that microbes play in the evolutionary development of multicellular organisms and also their impact on ecosystem functions (Berg et al. 2016).

1.4 Endophytes

Endophytes are microbes, that live inside plant tissues without causing any visible symptoms. Most plants, if not all, host microbial endophytes and they can be found in all plant parts and they are present in all stages of plant life (Frank, Guzmán, Shay 2017; Peñuelas and Terradas 2014; Truyens et al. 2015; Turner, James, Poole 2013). They may include commensal or mutualistic symbionts, but also latent pathogens. Endophytes can enter plant tissues in various ways, but two broad categories for endophyte acquisition can be defined: 1) Horizontal transmission where each plant generation acquires their endophytes anew from the environment, and 2) vertical transmission where the parental plant(s) transfers endophytes directly to the new seedling via seeds (Frank, Guzmán, Shay 2017).

In this study I use the definition of endophytes as microbes that reside in the interior of plant tissue without causing any visible symptoms, although some slightly differing definitions have also been proposed (Hallmann et al. 1997). Notably, this definition does not take into account the possible contribution the microbes have on the viability of the plant (i.e. whether the microbe is growth-promoting or not). It includes both mutualists and latent pathogens and anything in between. More specifically, in this study, unless otherwise stated, the term endophyte refers to bacterial leaf endophytes. Reference to other plant parts and microbial taxa will be given when needed.

1.4.1 Bacterial endophytes

Remarkable diversity of bacteria has been observed as plant endophytes and in many cases the nature of the interaction is well characterized (Hardoim et al. 2012; Hardoim et al. 2015; Hironobu and Morisaki 2007; Izumi et al. 2008; Miliute et al. 2015; Nissinen, Mannisto, van Elsas 2012; Sturz et al. 1997; Vega et al. 2005; Verstraete et al. 2011). Bacterial endophytes are best known for the

ability to fix atmospheric nitrogen and hence promote plant growth (Masson-Boivin et al. 2009). However, bacterial endophytes are also prominent plant pathogens (Strange and Scott 2005), as well as irreplaceable mutualistic symbionts (Hardoim, van Overbeek, van Elsas 2008; Santoyo et al. 2016). Previously the study of endophytic bacteria relied on culture-dependent methods where bacteria from internal plant tissues were isolated, cultured and identified. An obvious limitation of this method is, that it only captures the culturable part of the endophyte community (Hallmann et al. 1997). Development of sequencing techniques has allowed for culture-independent analysis of bacterial communities both outside and inside of plant tissues, which have confirmed that the culturable bacteria represent only a portion of the total plant-associated bacterial diversity (Berlec 2012; Izumi et al. 2008).

Many bacterial taxa have been observed commonly as endophytes. According to both culture-dependent and culture-independent studies, the most important groups of endophytic bacteria are Proteobacteria, Firmicutes, Bacterioidetes and Actinobacteria (Afzal et al. 2019; Araújo et al. 2002; Ding and Melcher 2016; Kumar et al. 2017; Rosenblueth and Martínez-Romero 2006). Endophyte community is of dynamic nature and is affected by plant species, season, climate, nutritional as well as developmental status of the host plant. Plant species have been shown to harbour specific bacterial communities (Ding and Melcher 2016; Hallmann et al. 1997; Kumar et al. 2017). This being said, it also seems, that most bacteria are in fact generalists when it comes to host plant species. Furthermore, bacteria are rarely observed as obligate endophytes but can instead shift from free-living stage to endophytic (Kumar et al. 2017). Season may play an important role in endophyte dynamics, especially for annual plants, where endophyte community is formed anew each year. Bacterial community has been observed to change over the course of seedling development (Barret et al. 2015). This seems logical since we know that soil bacteria play an important role in the endophyte community formation. Even if a portion of endophytes is transmitted from parents to offspring via seed, the soil bacteria have a big impact on the plant during germination.

1.4.2 Endophyte transmission and host control

Since soil-roots interface, also termed the rhizosphere, is the most important place where plant is in direct contact with microbes, it seems logical to assume that horizontal transmission from the surrounding soil is a significant, if not the most important, way how endophytes colonize plants. This has been confirmed in studies of root colonising endophytes, of which many are soil microbes entering the interior of roots (Compant, Clément, Sessitsch 2010; Compant et al. 2016). Plants are capable of choosing certain microbes in their roots (Jones et al. 2019). Plants can e.g.

produce exudates into the soil surrounding the roots which promote the growth of beneficial microbes. Bacteria also use chemotaxis to navigate towards roots where they may the establish symbiosis with the plant by remaining in the rhizosphere or by entering the root surface (rhizoplane) or the root interior (endosphere). From the root, bacteria can further spread systemically to other plant organs (Compant, Clément, Sessitsch 2010; Hallmann et al. 1997). Root nodule forming Rhizobium bacteria also enter the root from the surrounding soil (Sturz et al. 1997). The colonization process is overall tightly controlled by the plant. Plant deploys its defence mechanisms not only to deter pathogens but to selectively modify its microbiome (Jones et al. 2019).

From an evolutionary perspective, a most intriguing aspect is the possibility of vertical transmission of endophytes from parents of offspring. Vertical transmission could promote the establishment of tight mutualistic symbiosis and formation of dependency between the partners (Leigh 2010). An interesting example of a fairly recent establishment of strictly dependant and at least partly vertically inherited symbiosis has been observed between *Burkholderia* bacteria and several plant species in the Rubiaceae family (Pinto-Carbó et al. 2016; Verstraete et al. 2011). These bacteria form so called leaf nodules visible to the eye on the leaf surface as small dots. This symbiosis is so tight that the plant cannot live without its bacterial symbionts. The bacteria on some of the leaf nodule forming Rubiaceae species are also identified as novel species unable to survive outside the host (Carlier et al. 2016). Phylogenetic studies have shown that some coevolution has taken place, although no strict cospeciation, and horizontal endophyte changes has also happened (Lemaire et al. 2011). *Burkholderia* bacteria in Rubiaceae plants have been suggested to be involved in protecting the plant from herbivores via synthesis of toxic secondary metabolites (Sieber et al. 2015; Van Elst et al. 2013). However, genetic studies have shown that the genes responsible for synthesis of these metabolites are not present in all *Burkholderia* symbionts suggesting that it is not the sole reason for the intimate nature of the symbiosis. Several other bacterial symbionts not forming leaf nodules have also been observed from many Rubiaceae plants, some of which are closely related to soil bacteria (Lemaire et al. 2012). Although this example may be an exception rather than the rule when it comes to the specificity of the symbiosis compared to the vast diversity of other plant endophytic bacteria, it provides an interesting example of evolution towards a closed mutualistic and vertically inherited symbiosis.

Bacteria have been observed in seeds of many different plant species suggesting at least the possibility or some degree of vertical transmission. Vertical transmission of bacterial endophytes has been suggested e.g. to Norway spruce (Cankar et al. 2005), Marama bean (Chimwamurombe et al. 2016), several *Eucalyptus* species (Ferreira et al. 2008), Miscanthus (Cope-Selby et al. 2017)

and two rock-colonizing cactus species (Lopez, Bashan, Bacilio 2011; Puente, Li, Bashan 2009). In most of these cases confirming strict vertical transmission would require further analysis. Even more importantly, even if evidence of vertical transmission is found, its significance in shaping the final endophyte community can be questioned. At least several other factors play a role at the same time. Hardoim et al. (2012) found evidence of seed-borne rice endophytes being transmitted to the seedling, but also that soil properties including water conditions and pH played a crucial role in the formation of the endophyte community of the growing seedling.

1.4.3 Ecological significance

Rhizosphere microbes are essential in plant growth and nutrition: mycorrhizal fungi and nitrogen fixing bacteria are only the best-known examples of how soil microbes promote plant growth (Mendes, Garbeva, Raaijmakers 2013; Philippot et al. 2013). Same is likely to be true for many endophytic bacteria. In the soil, some bacteria enter plant roots as endophytes, most notably the nitrogen-fixing Rhizobium bacteria. Since soil is also an important source of plant-pathogenic microbes, the rhizosphere microbiome plays a crucial role in protecting the plant from disease (Mendes, Garbeva, Raaijmakers 2013; Raaijmakers et al. 2009). Mutualistic microbes can protect plants from pathogens e.g. by directly competing with pathogenic microbes or by inducing plant immune responses (Pieterse et al. 2014).

A common mechanism through which bacterial endophytes influence plant metabolism is the production of phytohormones, e.g. auxin or ethylene (Bhattacharyya and Jha 2012; Spaepen and Vanderleyden 2011). Auxin production is a typical trait of root-associated endophytes and it is also linked to enhanced endophyte colonization capacity. Auxins act as signalling molecules for both microorganisms and plants and they have a crucial role in plant defence responses. By modulating auxin levels in plants, microbes can either enhance plant defence responses or make plants more susceptible to pathogens. As noted earlier, endophytic lifestyle is a continuum from pathogenicity to mutualism, some endophytes being possibly latent pathogens. Phytohormone production by microbes is another example where there is no clear border between harmful and beneficial interaction: even difference between a toxin and a growth promoting hormone is not always clear since many pathogenic microbes use same pathways of stimulating plant metabolism as growth-promoting bacteria (Hardoim et al. 2015; López, Bannenberg, Castresana 2008).

Endophytic bacteria have also been linked to production of secondary metabolites in plants (Brader et al. 2014). Secondary metabolites typically play a role in defence and signalling. In the leaf-nodule symbiosis of *Burkholderia* bacteria and Rubiaceae plants, the bacterial symbionts seem to protect the plants from herbivores by producing secondary metabolites with insecticidal

properties (Sieber et al. 2015). In some Rubiaceae plants, the metabolites produced at least partly by the endophytic bacteria have also been linked to a fatal heart poisoning of ruminants (Van Elst et al. 2013; Verstraete et al. 2011). In these cases, the endophytic genomes do not exhibit the typical endophytic properties like phytohormone production, but they do harbour genes involved in secondary metabolite production, suggesting that this may be the key function of the symbiosis (Carlier et al. 2016). In another study, endophytic actinobacteria were found to stimulate the production of artemisin in herb *Artemisia annua* (Li et al. 2012). Artemisin is a potent secondary metabolite that is used for malaria treatment and hence the possibilities of bioengineering for its enhanced production are of great interest. In this study the genes involved in artemisin production were observed to be up-regulated after inoculation of the plants with endophytic bacteria. An open question is whether the bacteria is directly involved in the synthesis of artemisin or whether the observed up-regulation of the genes is a combined effect of the plant and the endophytes. These results also suggest that the endophyte promotes plant life by enhancing plant defences.

Endophytic bacteria are known to promote plant life by enhancing nutrient uptake e.g. by nitrogen fixation and phosphate solubilization (Afzal et al. 2019; Bhattacharyya and Jha 2012; Hardoim et al. 2015; Santoyo et al. 2016). They can furthermore promote plant life in harsh environments by modifying soil properties like nutrient availability and pH. In two cacti species, *Mamillaria fraileana* and *Pachycereus pringlei*, that colonize bare rock surfaces in dry desert areas, endophytic bacteria have been shown to enhance rock-weathering by modifying the pH on the rock surface and increasing nutrient availability (Lopez, Bashan, Bacilio 2011; Puente, Li, Bashan 2009). These bacteria have also been found in the surface-sterilized seeds of the cacti, suggesting vertical inheritance of endophytes. This kind of symbiosis where the symbiont enhances, or is maybe even a prerequisite for, the plant's colonization of novel ecological niches, can result in major ecological innovations, analogous to plants' adaptation to terrestrial life with the help of symbiotic relationships with fungi (Selosse and Le Tacon 1998).

1.5 Aim of the study

This study examines endophytes of *Plantago lanceolata*, the effect of soil on the endophytes and their possible vertical inheritance from parents to offspring. Since the bacterial endophytes of *P. lanceolate* have not been previously studied, the first aim of this study is to screen for bacterial leaf endophytes in *P. lanceolate* to get a comprehensive picture of the endophyte community. Although bacterial endophytes of *P. lanceolata* have not been studied before, fungal endophytes have been observed: Tello et al. (2014) reported *Hygrocybe virginea* as a systemic endophyte of

P. lanceolata. Hodgson et al. (2014) found 11 different fungal species as endophytes of *P. lanceolata*. Both studies detected these fungi in surface sterilized seeds and in seedlings grown in sterile conditions which suggests vertical transmission of endophytes from parent to offspring.

The second goal of this study is to assess the effect of maternal genotype on endophyte community composition by comparing the endophyte communities of offspring of different plant individuals that have been grown in similar conditions. At the same time, I will test how the soil affects the germination and growth of *P. lanceolata* seedlings. The growth experiment that I will use to study endophytes also enables testing how different soil treatments affect germination and growth. One set of seeds will grow in sterilized sand and the other in an environmental mixture of sand and soil collected from field conditions. This may give hints of the role of natural soil microbiota and nutrient conditions on germination and growth of *P. lanceolata* seedlings. This information can be especially useful in future studies on similar topics that involve germinating seeds in different conditions. Therefore, this test will also serve as method validation and development. Finally, I will try to elucidate the route of transmission of endophytes, if evidence of vertical transmission of endophytes can be found and how important the soil microbiota is in shaping the endophyte community.

Studies of bacterial endophytes have to date focused strongly on root endophytes of crop plants. Foliar endophytes and wild plants have received (quite understandably) less attention. However, the study of such systems could provide interesting and fruitful information on the nature of interactions between plants and bacterial endophytes. Since the bacterial density inside leaves is in general remarkably lower than in roots (Hallmann et al. 1997), they might be easier systems to study and infer knowledge on bacterial colonization and effect that they have on the plant.

This study will hopefully improve our understanding of the microbiome of *P. lanceolata* and potentially also more generally of the dynamics of plant (foliar) microbiome and the effect of soil in shaping endophyte community. Important new information would be better understanding of the effect of plant genotype on the endophyte community. We know for instance that plants select to some extent their endophytes from the surrounding soil and that endophyte communities are in some cases species specific. But intraspecific differences between individuals and genotypic differences in shaping endophyte communities are not known. Since the nature of the study is pioneering and rather explorative, it will hopefully give hints of possible future research directions.

2 Material and Methods

2.1 *Plantago lanceolata* & Åland study system

Plantago lanceolata (Plantaginaceae) is a widespread wind-pollinated perennial herb native to Eurasia. It reproduces both sexually, producing even hundreds of seeds annually, and clonally via side rosettes. In Finland its distribution is limited to the southern coast including the Åland islands and the rest of the southwestern archipelago. A typical habitat of *P. lanceoalata* in the Åland islands is a dry meadow.

P. lanceolata has been intensively studied in the Åland islands. It forms a network of several thousand populations which have been surveyed since 1991 as part of a long-term study of the Glanville fritillary butterfly (Hanski 2011; Ojanen et al. 2013). This network forms an ideal set up for metapopulation studies. The populations are spread out on an area spanning some 50 km by 70 km. The total land area of Åland islands is 1554 km^2. This set up has been successfully used for studying metapopulation dynamics of butterflies feeding on the *P. lanceolata* (Ojanen et al. 2013) and a fungal pathogen of *P. lanceolata* (Jousimo et al. 2014; Laine and Hanski 2006).

The plant material for the study was obtained from individuals grown in wild and from a growth experiment conducted in growth chambers in Helsinki.

2.2 Experimental Design

To study *P. lanceolata* endophytes, the following experiment was designed: Plants from a wild population were chosen to be sampled. These are hereafter referred to as the mother plants or maternal plants. The idea was to compare the leaf endophyte community of the mothers to that of their offspring. Seeds from these plants were used to conduct a growth experiment in laboratory to obtain leaf material from the offspring. Bacterial DNA was analysed from all leaf samples (mother plants and offspring) and endophyte community studied (figure 1).

2.3 Maternal plants from the wild

In June 2017 a total of 21 plant individuals were marked in a single *P. lanceolata* population in Järsö, Southern Åland (Picture 1). Population ID in the database is 332. Coordinates of the centre of the population are 6674735.682 N, 109413.447 E (ETRS-TM35FIN). Total area of the population is 1.72 hectares and *P. lanceolata* coverage is 15 m^2 (September 2017 survey). Three leaf samples were collected from each of the 21 individuals. The minimum distance between sampled plants was three meters. The leaves were placed in plastic tubes, stored in a cool bag and transported to Helsinki for further processing. In August 2017 all seeds and three more leaves of the same individuals were collected. All seed pods (if the plant had produced any) were collected and placed in a paper bag and transported to Helsinki. Furthermore, during the August sampling, soil was collected to be used in the growth experiment. Soil was collected from several different locations within the same population area, using a garden shovel, but not from the direct vicinity of the sampled plants. Soil was placed in 1 litre plastic bags and transported to Helsinki and stored in fridge temperature until the start of the growth experiment. For the experiment, all soil samples were mixed, not kept separately according to the exact collection spot, to form a uniform mixture.

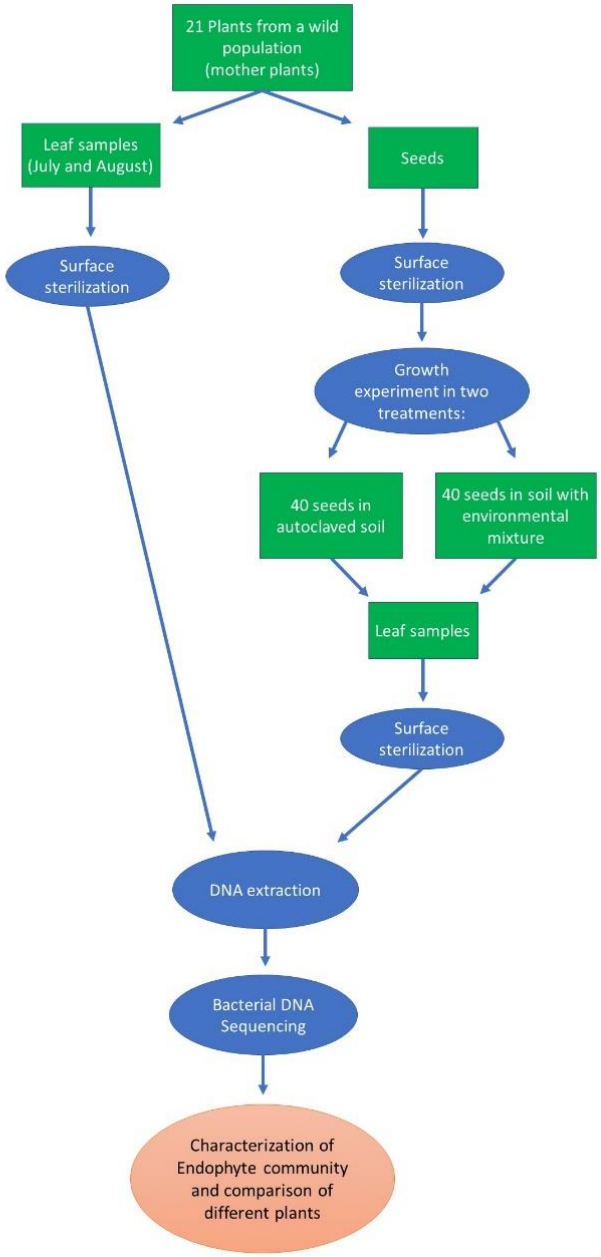
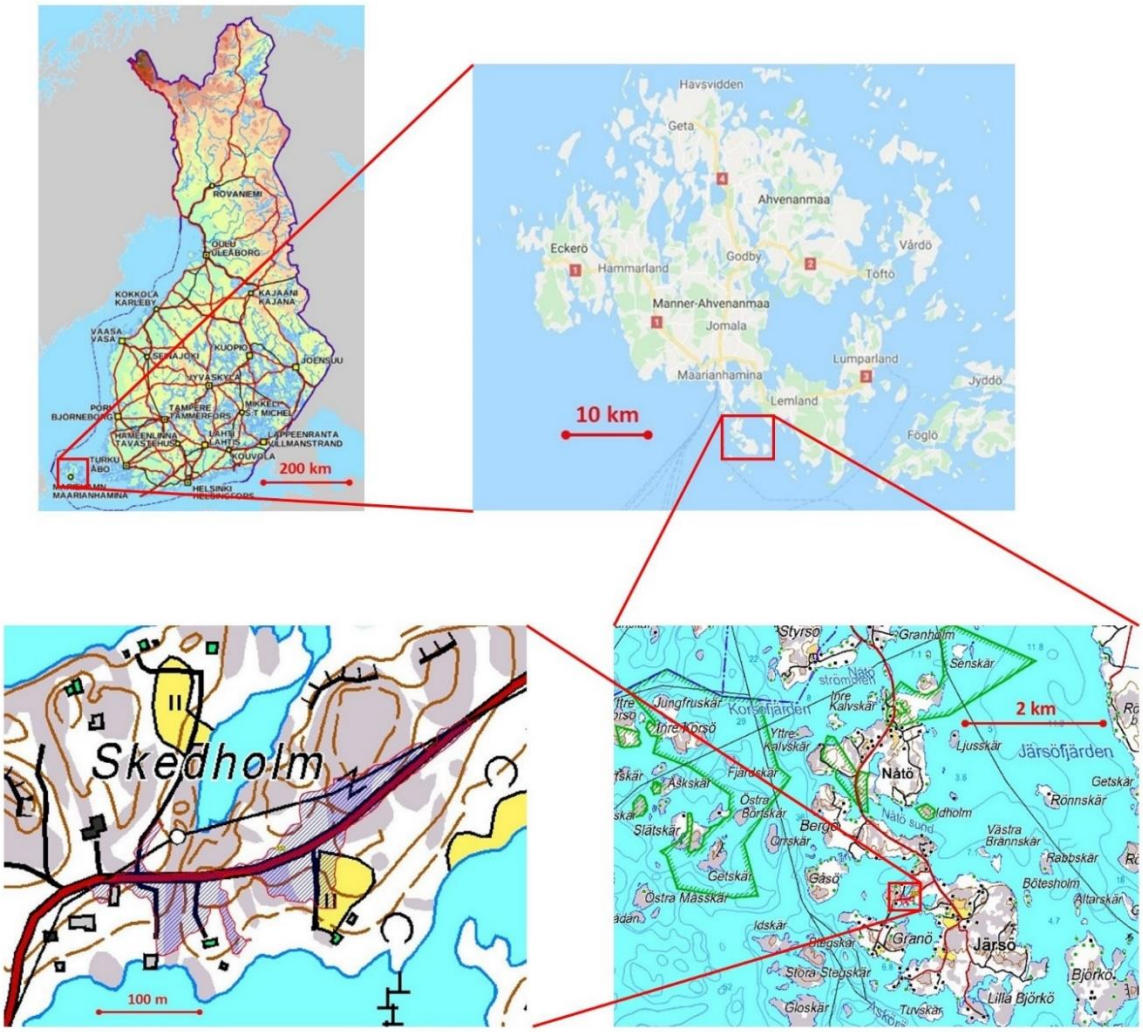


Figure 1. Diagram of experimental setup and different steps involved



Picture 1. Maps showing the location of the field site. *Top left: Finland. Top right: Åland islands. Bottom right: Järsö and surroundings. Bottom left: Field site in Järsö (P. lanceolata population marked with blue shade).*

Picture 2. Sampled individual of Plantago lanceolata in Järsö, Åland. *Plants were marked in June so that they could be found again in August when the second sampling took place.*

2.4 Sterilization and sampling

In Helsinki the leaves were surface sterilized within 36 hours from collection to eliminate any epiphytic microorganisms. The sterilization was executed by submerging the whole leaf into 70% ethanol for one minute, 2.7% chlorite for three minutes and finally rinsed three times one minute in sterile double distilled water (Picture 4). Rinsing water was changed after 5-7 samples. Sterilized leaves were placed on a paper towel to dry for a few minutes. Then samples of approximately 100 mg were cut from each leaf, placed in 2 ml microcentrifuge tubes and stored in -80^{o}C to wait for DNA extraction.

Picture 3. Leaf sterilization in action

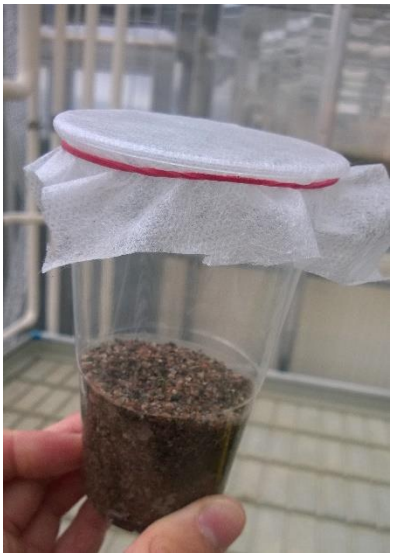
Picture 4. Sterilization setup*: leaves were transferred in metal tea bags and submerged (from left to right) in 1) 70% ethanol, 2) 2.7% chlorite and 3)-5) H_20.*

Same surface sterilization protocol was applied for seeds just before the start of the growth experiment. The success of sterilization was tested by taking samples of the last rinsing water after the sterilization of 5-7 samples was completed. 100 µl of this water in three replicates was poured on a growth medium and left to incubate in room temperature for two weeks. No microbial growth of any kind was observed.

2.5 Growth Experiment

In the growth experiment, seeds from different individuals (genotypes) were grown in two different treatments: 1) in sterilized (twice autoclaved) sand and 2) in a mixture of sterilized sand and soil collected from the wild (from now on referred to as "environmental mixture"). The environmental mixture was intended to mimic the biological properties, most importantly the microbiome, of the soil where the plants normally grow. Sterilized soil on the other hand would be virtually lifeless and the seeds germinating in it would not confront any of the normal soil microbes. The sand used was autoclaved twice with three days in between to better get rid of any microbes, including possible fungal resting spores. The success of the sterilization was tested by scattering one spatula of sand on a growth medium that was left to incubate for two weeks. Some fungal growth could be observed but no visible bacterial growth. Environmental mixture was prepared by mixing three parts of autoclaved soil with one part of soil collected from the wild. Mixing was done by hand with a large bucket and a shovel.

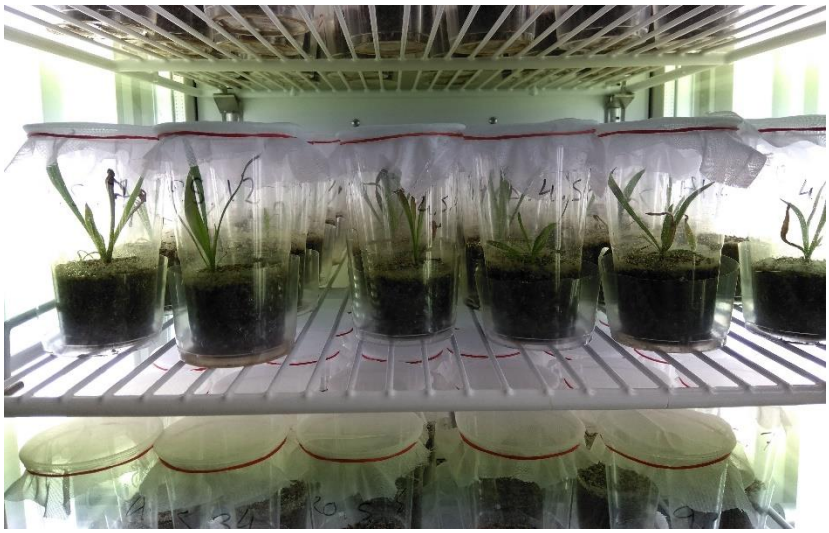
Growth experiment was started in the end of September 2018 in Viikki, Helsinki. Out of the 21 plants sampled, five were selected to be used in the growth experiment. These five plants were among those that had produced sufficient amount of seeds needed for the experiment. They will be henceforth referred to as the mother plants. 80 seeds of each mother plant were sown. In addition, 60 seeds from an extra mother plant were sown in sterile soil to be used as a control for the growth chambers.



Picture 5. Growth pot with sterilized soil

All seeds were surface sterilized just before sowing as described above. 40 seeds from each mother plant were sown in sterilized soil and 40 in environmental mixture, each in its own plastic growth pot kept individually on their own plastic plates (Pictures 5 and 6). This way the mixing of soils and movement of microbes between pots was minimised. Sowing was performed in a microbe laminar. Each pot was covered with a piece of gauze to prevent any insects from getting in contact with the plants. Pots with sterilized soil were then placed in one growth chamber and pots with environmental mixture to another, again to minimize any cross-contamination of microbes from the environmental mixture to the sterile soil. To monitor any possible growth chamber effect, additional seeds from one extra maternal plant were sown, three seeds per pot,

in sterilized soil and these pots were placed in the same growth chamber as the pots with environmental mixture.



Picture 6. Growth pots in growth chamber. Each pot is kept on their individual plastic plate.

Pots were left to incubate in the growth chambers with a diurnal pattern of 16 hours of light, +21°C and 8 hours of darkness, +15°C. Plants were watered once a week with autoclaved tap water. Five weeks after the start of the experiment, additional fertilizing with NEKO Plant Fertilizer (N-P-K 7-2-2) was started to both treatments. To minimise any transmission of microbes with the fertilizer, the fertilizer was filtered before use through a microfilter (pore size 45 µm). The success of the filtration was tested by pouring 100 µl of filtered fertilizer on a growth medium that was left to incubate in room temperature for two weeks. No microbial growth was observed.

The number and size of leaves in each plant were measured every week during the growth experiment and the pots´ location within the growth chamber rotated. After seven weeks the experiment was terminated and samples from all plants collected. All remaining leaves were collected and surface sterilized in the same way as the original material from the wild (described above). Pieces of approximately 100 mg were cut from the leaves, placed in microcentrifuge tubes and stored in -80^{o}C. If leaves were too small so that there was not enough leaf material to obtain a sample of 100 mg, the whole leaf (or several) was placed in the microcentrifuge tube to get as much material as possible.

2.6 DNA extraction

DNA extraction was done in March-April 2018 in Helsinki using Qiagen Plant DNA extraction kit. Leaf samples of approximately 100 mg were manually ground in Eppendorf microcentrifuge tubes using liquid nitrogen. Grinding proceeded in sets of four to six samples. All samples of one set were frozen and ground and extraction buffer 1 added to them before grinding the next set of samples. DNA was extracted from one leaf sample per plant of all plants grown during the growth experiment and the six maternal plants, sampled both in June and August (five for the actual experiment and one for growth chamber control). Extracted DNA was stored in 50 µl of double distilled water and frozen in -20°C to wait for sequencing. A total of 11 empty controls were done on separate days, with the same extraction steps as in the real samples but without any material in the sample tubes.

2.7 16S rRNA

To analyse endophytic bacteria, culture-independent methods need to be applied. Some endophytes may be culturable outside the host, but a significant portion of the endophytic diversity will remain hidden if only culturable bacteria are assessed. To reveal this hidden diversity, endophytic DNA must be isolated and analysed.

The most widely used method for bacterial taxonomy, is the analysis of 16S rRNA gene (Baker, Smith, Cowan 2003; Ghyselinck et al. 2013; Yarza et al. 2014). 16S rRNA is a part of the small sub-unit of prokaryotic ribosomes. 16S rDNA refers to the gene responsible for coding this RNA molecule. Since ribosomes are essential components of any living cells, mutations to genes encompassing their functioning would generally be detrimental to the organism. Because of this, ribosomes' structure and function have been highly conserved over evolution and they are minimally affected by lateral gene transfer (Head, Saunders, Pickup 1998). Most importantly for bacterial taxonomy, the 16S rRNA exhibits high level of conservation. It is some 1500 bp long and includes regions that are totally or mostly conserved as well as variable and highly variable regions (Baker, Smith, Cowan 2003). This makes it possible to infer taxonomic relationships between different bacterial strains. The more conserved areas of 16S region allow for a reliable alignment and comparison of DNA strands, whereas the dissimilarities in the less conserved areas allow estimation of taxonomic distance between the strains.

2.8 PCR & DNA sequencing

Before the actual PCR and sequencing, a preliminary test run was performed in Viikki, Helsinki, to validate the methods. This preliminary PCR amplification and sequencing was performed on three samples. After the preliminary test, DNA samples were sent to the Institute of Biotechnology in Helsinki for the actual PCR amplification and DNA sequencing.

Since the eukaryotic chloroplasts and mitochondria are of prokaryotic origin and harbour genes of their own, they also have 16S rDNA regions, which can easily interfere with the analysis. To prevent this, it was necessary to use primers that would amplify the bacterial sequences but exclude any sequences from the plant mitochondria or chloroplasts. Therefore, the PCR was done with bacterial 16S rDNA primers designed originally by Chelius and Triplett (2001). These primers would exclude nearly all chloroplast sequences. Mitochondria sequences are however not as easily separable and further processing of PCR products is necessary to exclude mitochondrial reads (described below).

The initial PCR amplification was done using primers 799f, 5'AACMGGATTAGATACCCKG'3 and 1492r (Chelius and Triplett 2001). Since the resulting DNA products would have been too large for Illumina sequencing, a nested PCR was performed with primers 1062f and 1390r (Ghyselinck et al. 2013; Zheng et al. 1996), using the products of the first PCR as templates (Kumar et al. 2017). These primers target the V6-V8 region of the 16S rDNA and produce amplicons with variable regions resulting in high phylogenetic coverage (Baker, Smith, Cowan 2003) and proper size for Illumina sequencing. The PCR reactions contained 2 µl of template DNA, 5x HF Buffer, 10 mM dNTPs, 10 µM of each primer, 2 U/µL Phusion HS II DNA Polymerase and 0.625 µl DMSO 100% in a 25 µl reaction volume. First PCR amplification was performed as follows: 30 seconds denaturation at 98°C followed by 20 cycles of denaturing DNA strands, annealing primers, and extension of new strands at 98°C for 10 seconds, 55°C for 30 seconds, and 72°C for 10 seconds, respectively, with final extension lasting for 5 minutes. The nested PCR amplification was performed in a similar way except that only 14 rounds of denaturing, annealing and extension were used.

Mitochondrial DNA products were supposed to be separated from bacterial products by size fractionation. According to Chelius and Triplett (2001), use of the aforementioned primers would result in mitochondrial PCR products that are bigger in size compared to the bacterial ones. This would enable the separation of the endophyte products from the mitochondrial products by running them on an agarose gel by electrophoresis which should then show two distinct bands: mitochondrial and bacterial DNA products separately. Kumar et al. (2017) and Nissinen, Mannisto,

van Elsas (2012) successfully used this approach to amplify endophyte DNA. After our preliminary test PCR amplification, DNA products were run on gel electrophoresis to separate the products of different size. However, no clearly separate bands were distinguished on the gel, so it was decided not to do the size fractionation. Samples were then sent to the Institute of Biotechnology in Helsinki for the PCR amplification and Illumina next-generation sequencing. Four empty controls were also run with no samples in them. Sequencing was done twice. First run resulted in over 11 million reads and the second one in over 6 million reads. To get a better coverage and detection probability of endophyte sequences, sequences from the first run were decided to be used.

2.9 Sequence processing & OTU clustering

I processed the sequence reads using CLC Genomics Workbench 11.0 with Microbial Genomics Module (Qiagen.com). I first screened the raw reads imported from Illumina (XX) for chimeras, and removed reads less than 150 base-pairs and with Q score less than 25. After removing bad quality reads, 3,484.464 reads remained for the analysis. These good quality reads I clustered into OTUs (Operational Taxonomic Units) at 97% sequence identity, and then assigned taxonomy for the OTUs using SILVA reference database version 137.

2.10 Statistical & Community Analysis

To test the effect of treatment and maternal genotype on germination and plant size, I fit two generalized linear mixed models where germination (0/1) was defined as a binary variable with a logit link function, and length of longest leaf as continuous variable with a normal distribution of errors as visually verified from data. In both models treatment and maternal genotype were defined as categorical variables, and sample was defined as a random effect nested under maternal genotype. If significant, interactions were retained in the model. The analyses were run in JMP 14.0.

To understand whether the bacterial communities in *P. lanceolata* leaves differed between genotypes and soil treatments, I performed statistical analysis of the bacterial communities. The analysis is based on dissimilarity values of different groups which are then used to perform Analysis of Similarities (ANOSIM). I analysed the bacterial communities as presence/absence data. This means that I did not account for the number of reads, and only used information on whether a bacterial genus had been identified from a certain sample or not. I then pooled samples according to the treatment in order to be able to compare bacterial communities of different treatments, so the unit of analysis was the treatment. I added together the observations of all

bacterial genera from all plants from a certain treatment. This resulted in the bacterial community data of the whole treatment, e.g. the bacterial community of the seedlings of mother plant #4 grown in sterile soil (treatment code 4.S). For the obtained community data, I calculated dissimilarity values using R software (version 3.5) package vegan. I used Bray-Curtis dissimilarity index which is defined as:

Bray-Curtis: $$\frac{\sum|x_i - x_j|}{\sum(x_i + x_j)}$$

Where x_i and x_j are the abundances of each species in groups i and j (Gardener 2014). In this case the abundance of each species means the number of plants in a certain treatment where a specific bacterial genus was observed. Therefore, x_i means number of plants in treatment i, where the bacterial genus x was found. Dissimilarity is a value between 0 and 1, and the higher the number is (i.e. closer to 1), the more different the two groups are. 0 therefore means identical communities, where there is no difference in species abundance and composition. I calculated dissimilarity values for all pairs of treatments, which resulted in a list of dissimilarity values, which tells us how similar the bacterial communities of any two treatments are to each other. Based on the Bray-Curtis dissimilarity index, I drew ordination plots to graphically illustrate the differences between communities. Furthermore, I tested Analysis of Similarities (ANOSIM, vegan 2.5.6) to assess the statistical difference between the communities. All analyses were done with R software version 3.5.

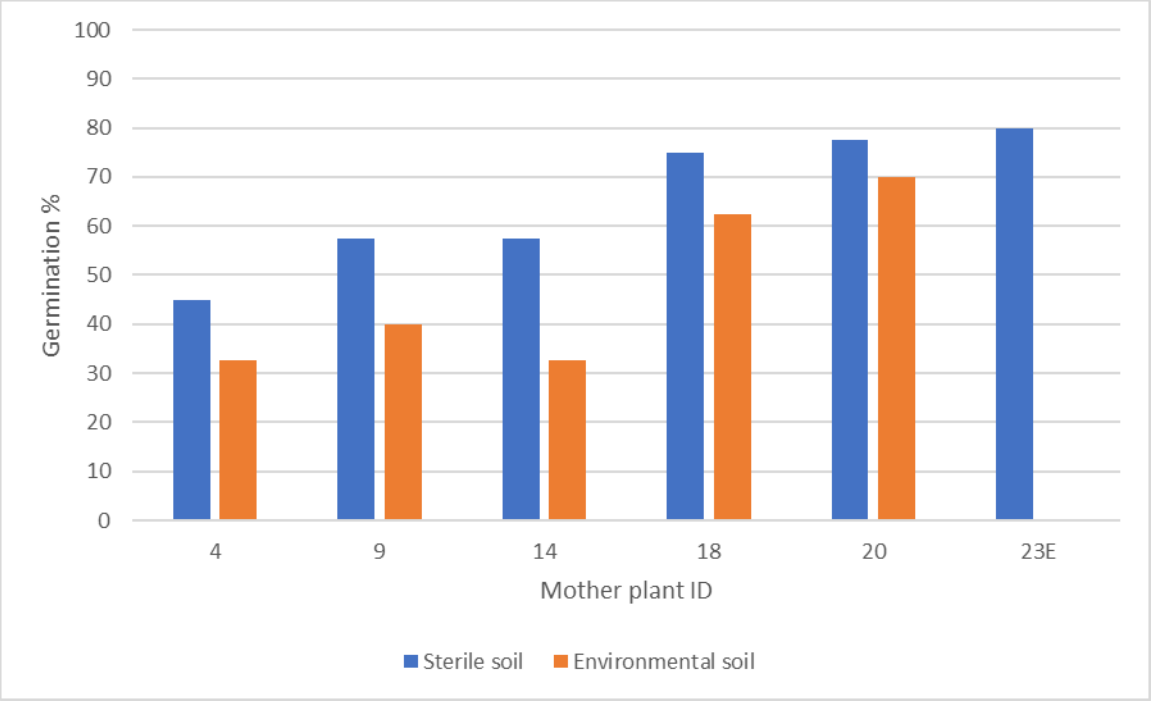
3 Results

3.1 Growth Experiment

From each of the five mother plants, I sowed 80 seeds, 40 in sterile soil and 40 in environmental soil. In addition, I sowed 60 seeds from an extra mother plant as growth chamber controls. Out of all the seeds sown, less than half germinated. Germination results are shown in Table 1 and Picture 7.

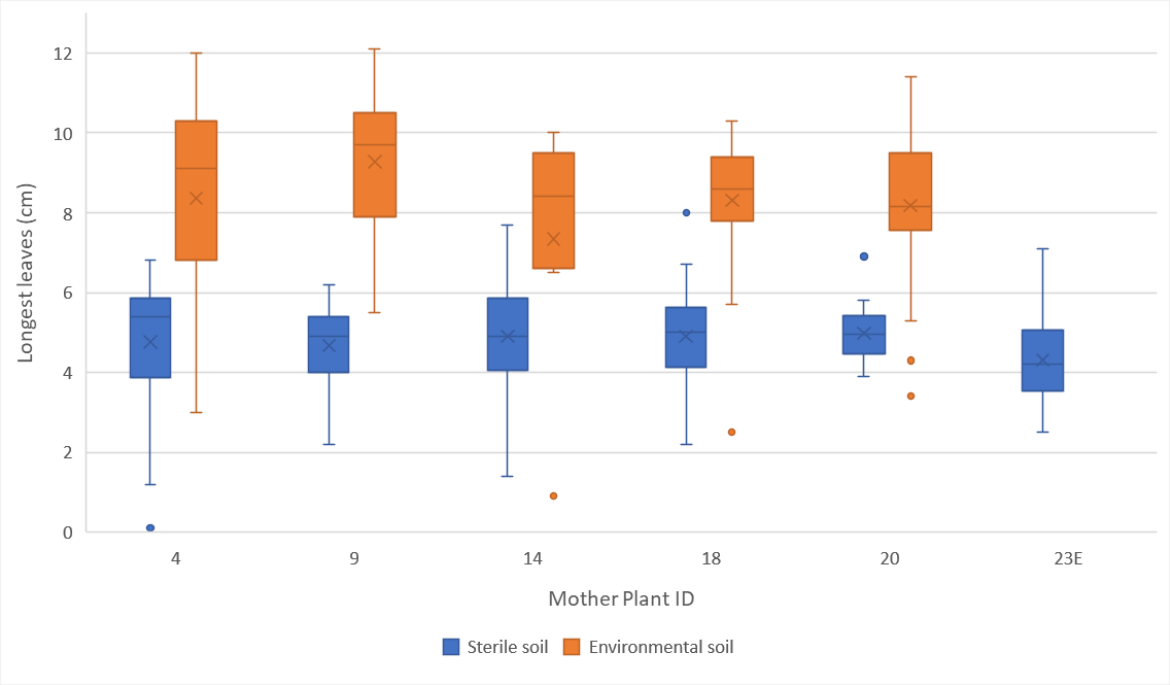
Table 1. Germination of seeds. *The columns are, respectively:*
Mother plant: the ID number of the mother plant; Treatment: Sterile = seeds grown in twice autoclaved sand, GC control = seeds grown in twice autoclaved soil but grown in growth chamber with Environ treatments, Environ = seeds grown in soil mixture of sand and soil collected from wild population in Åland; Seeds sown: number of seeds sown per treatment; Seedlings: number of seeds that germinated and grew into seedlings; Germination %: the percentage of the seeds that germinated; Final sample size: number of seedlings that survived until the end of the experiment so that leaf samples could be acquired.

Mother plant	Treatment	Seeds sown	Seedlings	Germination %	Final sample size
4	Sterile (S)	40	18	45.0%	18
9	Sterile (S)	40	23	57.5%	23
14	Sterile (S)	40	23	57.5%	22
18	Sterile (S)	40	30	75.0%	30
20	Sterile (S)	40	31	77.5%	30
23E	GC control	60	48	80.0%	48
4	Environ (Y)	40	13	32.5%	11
9	Environ (Y)	40	16	40.0%	15
14	Environ (Y)	40	13	32.5%	13
18	Environ (Y)	40	25	62.5%	25
20	Environ (Y)	40	28	70.0%	28



Picture 7. Germination percentage per treatment. *Blue bars: seedlings grown in sterile soil. Orange bars: seedlings grown in environmental mixture soil. In all groups, seeds grown in the environmental soil mixture had clearly a lower germination percentage.*

For all genotypes, germination percentage was significantly lower in the environmental soil than in the sterile soil (F = 10.78, *P* = 0.0012). Lowest germination percentage was in the seeds of mother plants #4 and #14 in the environmental soil (treatments 4.S and 14.S, germination % 32.5%) and highest in the growth chamber controls, which were grown in sterile soil (treatment 23E.S, germination % 80%). Genotype had also a significant effect on the germination (F = 7.98, *P* < 0.0001), with seeds from certain mother plants having clearly higher germination percentage in both sterile and environmental soil as compared to seeds of other mother plants. The interaction between treatment and maternal genotype was not statistically significant. Seedlings grown in environmental mixture ended up bigger than seedlings in sterile soil (Picture 8; F = 10.91, *P* < 0.0001). Maternal genotype did not have a significant effect on leaf length (F = 0.72, *P* = 0.5787), nor was there a significant interaction between treatment and maternal genotype.



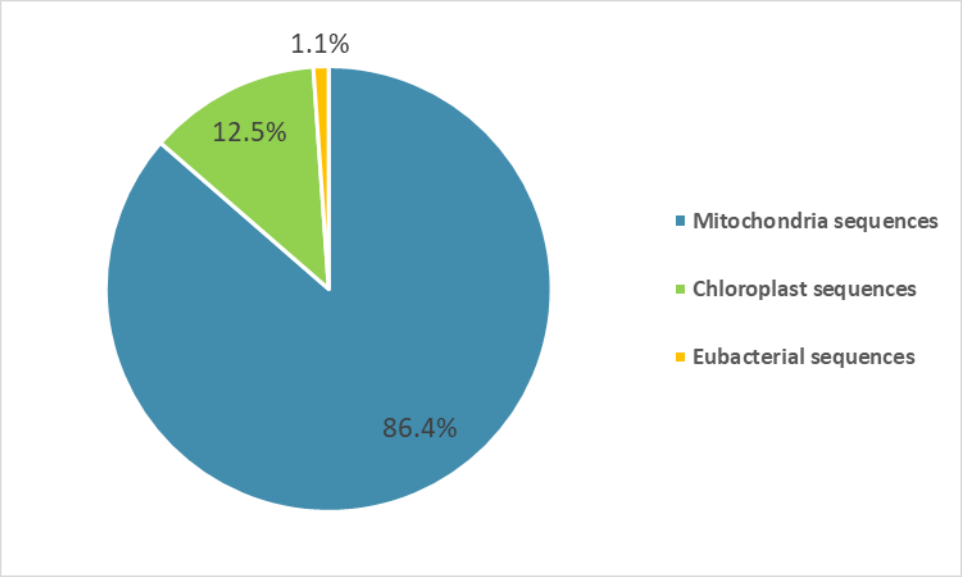
Picture 8. Boxplot of the lengths of the longest leaves at the end of the growth experiment. *Blue represents plants grown in sterilized soil, orange represents plants grown in environmental mixture. Number of the mother plant is shown on the horizontal axis, leaf length on the vertical axis.*

3.2 Bacterial diversity

After filtering chimeric and bad quality reads, 3,488,048 reads remained for OTU (Operational Taxonomic Unit) clustering. Vast majority of these turned out to be *P. lanceolata* mitochondria (3,013,451 reads) or chloroplast (436,116 reads) sequences. Only 38,481 reads, about 1.1% of the total, were of eubacterial origin (Table 2, Picture 9). When mitochondria and chloroplast reads were removed, the remaining reads clustered into 218 different bacterial OTUs. Most of these were identified only to the genus level.

Table 2. Number of reads obtained in three main categories

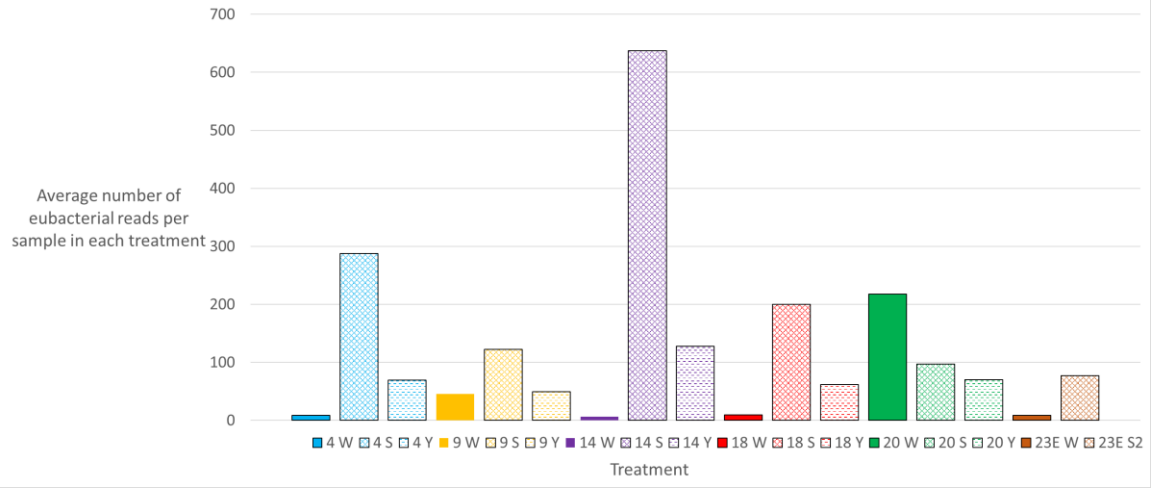
	No. of reads	% of total
Mitochondria sequences	3,013,451	86.4
Chloroplast sequences	436,116	12.5
Eubacterial sequences	38,481	1.1
Total sequences	3,488,048	100



Picture 9. Percentages of reads obtained from three different categories: *vast majority, 86.4%, were P. lanceolata mitochondria sequences, 12.5% were P. lanceolata chloroplast sequences and only 1.1% were eubacterial sequences.*

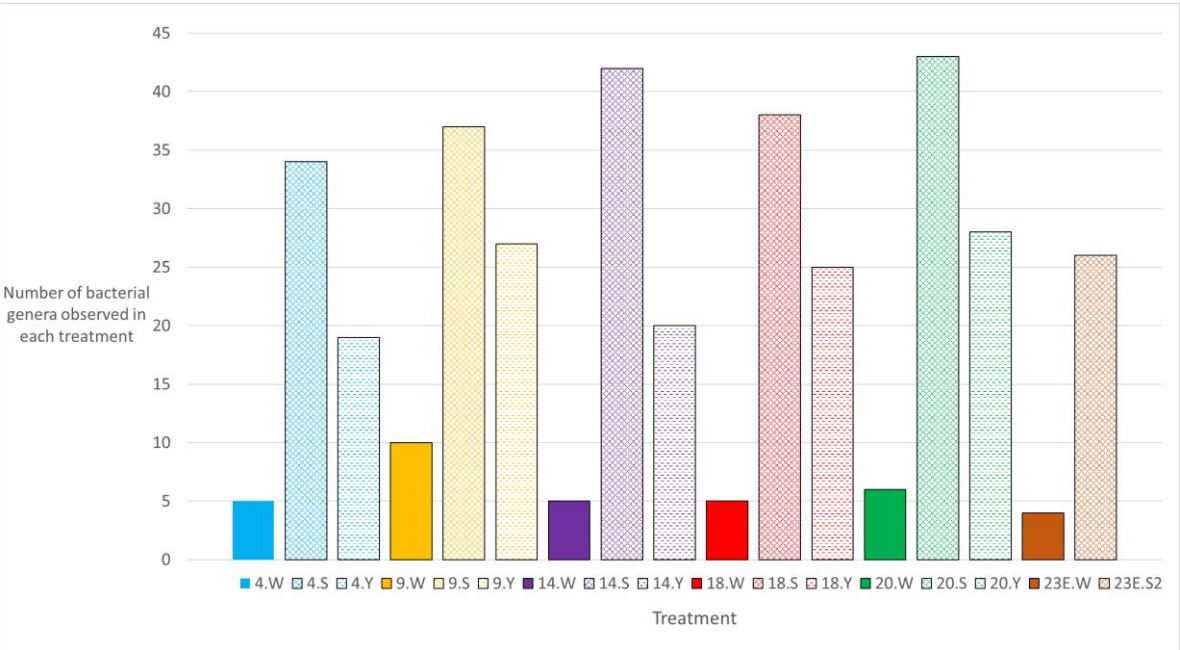
In total, I identified 71 eubacterial genera from the data. Six most common genera constituted 83.1% of all eubacterial reads and 0.9% of all reads (total of 31,981 reads). These were *Shewanella* (10,254 reads), *Methylobaterium* (8,205 reads), *Ralstonia* (6,045 reads), *Halomonas* (4,677 reads), *Salinicola* (1,733 reads) and *Clostridium* (1,067 reads) (Table 3). The remaining 16.9% of eubacterial reads (6,500 reads, 0.2% of all reads) clustered into 166 different OTUs, representing 65 different bacterial genera.

In general, the amount of eubacterial DNA reads obtained from different treatments follows the same pattern: samples from sterile soil treatment yielded highest number of eubacterial reads, samples from environmental mixture second highest and samples from mother plants lowest (Picture 10). The only exception is mother plant #20, which yielded more eubacterial reads per sample than its offspring. This is due to high numbers of reads from two specific bacterial OTUs found in mother plant #20 (*Candidatus* Nasuia deltocephalinicola and *Candidatus* Sulcia muelleri, which I will discuss later).



Picture 10. Average number of Eubacterial reads obtained from a single sample in different treatments. *Horizontal axis: number and the corresponding colour indicates the mother plant and its offspring, letters indicate the growth conditions during the experiment, W = Wild (mother plant), S = Sterile (seedlings grown in sterilized soil), Y = Environmental (seedlings grown in environmental mixture soil), S2 = Growth chamber control. In all groups, except #20, bacterial abundance follows the same pattern: S has highest number of reads, Y second highest and W lowest.*

Similar pattern continues when looking at the genus richness in different treatments: Sterile treatment group has the highest genus richness, environmental mixture group second highest and mother plants the lowest (Picture 11).



Picture 11. Number of bacterial genera found in each treatment. *Horizontal axis: number and the corresponding colour indicates the mother plant and its offspring, letters indicate the growth conditions during the experiment, W = Wild (mother plant), S = Sterile (seedlings grown in sterilized soil), Y = Environmental (seedlings grown in environmental mixture soil). Vertical axis shows how many different bacterial genera were observed from each treatment. Samples are pooled together into treatments to get a picture of the bacterial genus richness in each group.*

Table 3. Six most common bacterial genera in numbers. *Total sample size is 258. The columns are, respectively: Genus is the six most common bacterial genera; No. of samples where present says how many of the sequenced samples yielded at least one read of the genus (that genus is therefore considered present in the sample); % of samples where present says the percentage of the samples (out of the total of 258) where the genus is present; No. of reads says how many reads of this genus were obtained in total; % of eubacteria reads says what percentage of all the eubacterial reads represents this genus; % of total reads says what percentage of all the reads represents this genus.*

Genus	No. of samples where present	% of samples where present	No. of reads	% of eubacterial reads	% of total reads
Shewanella	255	98.80%	10,254	26.70%	0.30%
Methylobacterium	60	23.30%	8,205	21.40%	0.20%
Ralstonia	144	55.80%	6,045	15.70%	0.20%
Halomonas	256	99.20%	4,677	12.20%	0.10%
Salinicola	219	84.90%	1,733	4.50%	>0.1%
Clostridium	2	0.80%	1,067	2.80%	>0.1%
Total			31,981	83.20%	0.90%

3.2.1 Abundant bacterial genera

Below I will describe the occurrence of the six most common bacterial genera (Table 3) in more detail. In addition, I will describe the occurrence of two intriguing bacterial genera, *Candidatus Nasuia deltocephalinicola and Candidatus Sulcia muelleri*.

Shewanella

Shewanella was the most common detected bacterial genus. I observed *Shewanella* reads from 255 samples out of the total of 258 samples that yielded leaf material for extraction. In total I detected 10,254 DNA reads from *Shewanella* bacteria, representing 26.7% of all eubacterial reads (Table 3). *Shewanella* reads were also widely present in empty control samples, done both during DNA extraction and sequencing phase (Table 4). It seems highly likely that the observed *Shewanella* are contaminants.

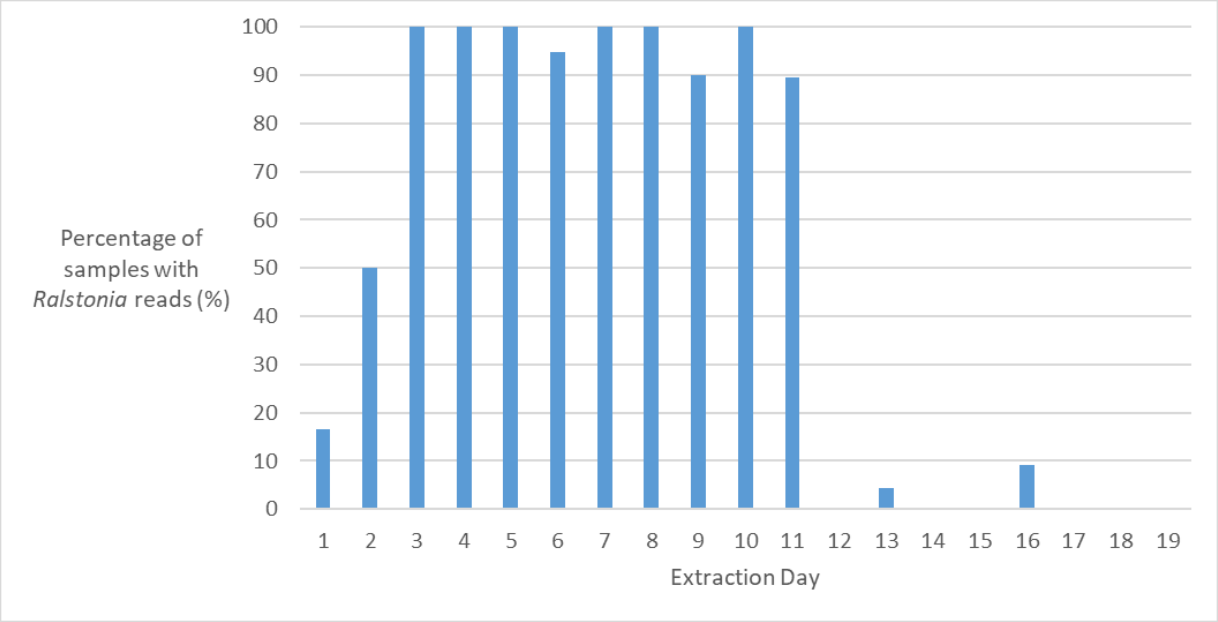
Methylobacterium

Methylobacterium is the second most abundant bacterial genus in the samples based on the number of DNA reads. There was a total of 8,205 reads representing *Methylobacterium* bacteria.

However, they only appear on 60 of the 258 samples (Table 3). Furthermore, only two samples yielded by far most of the *Methylobacterium* reads: sample #59 (treatment code 14.S.15) had 4,184 reads and sample #233 (treatment code 14.S.17) had 3,537 reads. This means, that 94.1% of all *Methyobacterium* reads came from only two samples. These sample represented two different plant individuals (ID numbers #15 and #17) but they came from the offspring of the same mother plant (ID number #14) grown in sterilized soil (treatment code S). The remaining 58 samples that harboured *Methylobacterium* reads had together only 484 reads, ranging from only 1 read per sample to 85 reads per sample. This kind of occurrence could indicate pathogenicity. However, *Methylobacterium* are not known as plant pathogens but as common phyllosphere symbionts (Knief et al. 2010). *Methylobacterium* reads were not present in control samples.

Ralstonia

Ralstonia is the third most common bacterial genus by the number of reads (total of 6045 reads). I did not observe *Ralstonia* reads in any controls, however, its occurrence virtually ends completely after the 12th extraction day. During the first 11 extraction days, *Ralstonia* occured in 91.6% of the samples (142 of the 155 samples extracted during that time). From the 12th day on, *Ralstonia* occurred in only 1.9% of the samples (2 of the 103 samples extracted during that time, Picture 12). This timing coincides with the changing of the DNA extraction kit. Until the 12th extraction day, I used an old and previously opened extraction kit and started a new one on the 12th extraction day, after the reagents in the first one run out. This raises doubt that *Ralstonia* reads could be contamination from the old extraction kit. Samples were not extracted in a specific sampling order, but different treatments and genotypes followed each other in a non-systematic manner. The only exception to this is, that samples from the mother plants were processed during the first two extraction days.



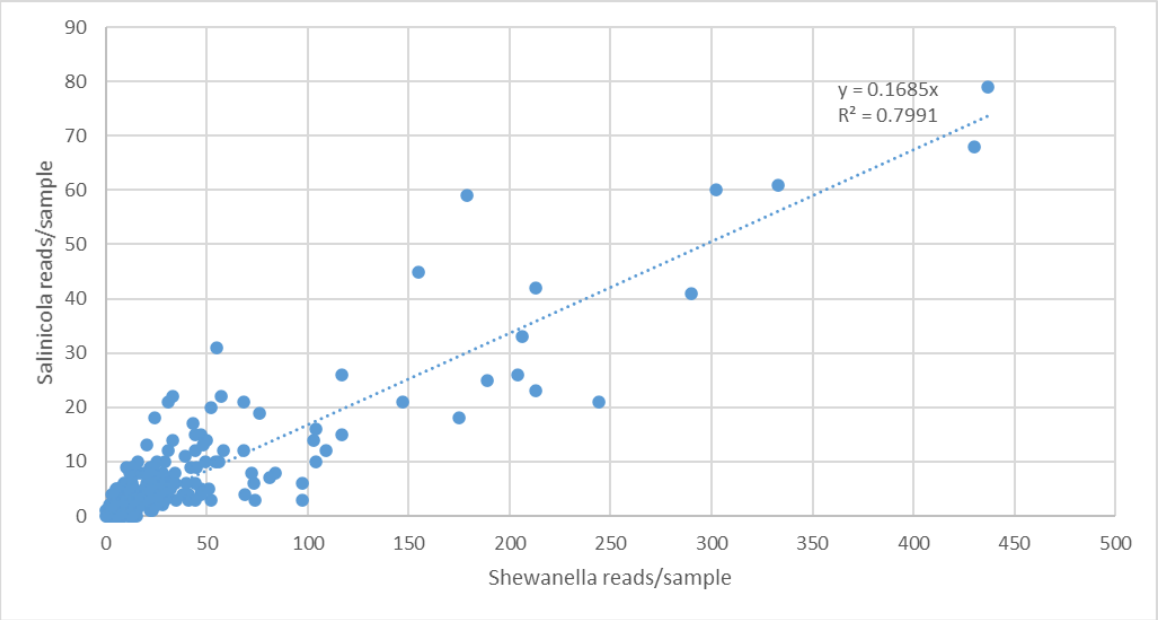
Picture 12. Percentage of samples containing Ralstonia reads on every DNA extraction day. *Picture shows that up until the 11th extraction day Ralstonia reads were obtained from almost all samples but they virtually disappear after that. Here Ralstonia observations are only considered as present/absent, ignoring the number of reads obtained. If any number of reads was obtained from a sample, the bacteria genus is considered present in the sample, if zero reads are obtained, the bacterial genus is considered absent.*

Halomonas

Halomonas is, after *Shewanella*, the second most common bacterial genus in empty control samples. Like *Shewanella*, I found *Halomonas* in controls done both during DNA extraction and during sequencing. I obtained *Halomonas* reads from all but one control sample. It therefore seems obvious that *Halomonas* reads are a result of contamination. In total, I obtained 4571 *Halomonas* reads. They were present in all groups: mother plants and sterile and environmental treatments as well as in growth chamber controls.

Salinicola

Salinicola reads are present in some controls, in both extraction and sequencing controls, but not in large numbers (Table 4). Still this presence in empty controls suggest contamination. Furthermore, there is strong correlation between the presence of *Shewanella* (a clear contaminant) in the samples and the presence of *Salinicola* in the samples (Picture 13). In total, I observed 1708 reads of *Salinicola*. They were present in all treatment groups: in mother plants, sterile and environmental treatments and growth chamber controls.



Picture 13. Correlation between Shewanella and Salinicola occurrence. *Every blue dot represents one sample. Horizontal axis gives the number of Shewanella reads in the sample. Vertical axis gives the number of Salinicola reads in the sample. The strong correlation with Shewanella, which is a clear contaminant, may suggest contaminant origin also for Salinicola.*

Clostridium

Clostridium reads were obtained from two samples only, A111 and A234 (also not in controls). These two samples present a curious exception in the growth experiment. They represent two separate plant individuals, that were by accident sown in the same growth pot. Both seeds came from mother plant number 14 and grew in environmental mixture soil (treatment code 14.Y). Normally only one seed was planted per pot, but in this case another seed ended up in the same growth pot apparently because of carelessness during sowing. One of the samples yielded 1065 *Clostridium* reads, the other one only two. These plants grew poorly and almost died during the experiment. The leaves were deemed to be “almost dead” by the sixth week of the experiment and in the seventh and final week of the experiment new small leaves had appeared from which the samples were taken. All this seems to point in the direction that *Clostridium* could be a pathogen and could have resulted in the poor performance of the seedlings.

Table 4. Six most common bacterial genera in control samples. *Type = control done during extraction/sequencing (Extr = Extraction, Seq = Sequencing); Numbers under each bacterial genus show the number of reads of the given genus obtained from each control sample.*

Control sample ID	**Type** (Extraction/ Sequencing)	***Shewanella***	***Methylobacterium***	***Ralstonia***	***Halomonas***	***Salinicola***	***Clostridium***
A152	Extr	3	0	0	2	0	0
A153	Extr	30	0	0	9	3	0
A172	Extr	24	0	0	8	1	0
A173	Extr	5	0	0	4	1	0
A195	Extr	46	0	0	21	6	0
A196	Extr	2	0	0	1	2	0
A216	Extr	2	0	0	2	0	0
A217	Extr	3	0	0	1	1	0
A243	Extr	5	0	0	1	0	0
A248	Extr	8	0	0	6	2	0
A252	Extr	4	0	0	2	2	0
A253	Seq	6	0	0	1	0	0
A254	Seq	0	0	0	1	0	0
A255	Seq	102	0	0	47	7	0
A256	Seq	0	0	0	0	0	0

Candidatus Nasuia deltocephalinicola and *Candidatus* Sulcia muelleri

I detected two species of bacteria, *Candidatus* Nasuia deltocephalinicola and *Candidatus* Sulcia muelleri, that are known as obligate gut symbionts of sap-feeding insects (Bennett and Moran 2013; McCutcheon and Moran 2010; Noda et al. 2012). I detected them from two samples (A003 and A224), which came from leaves of one of the mother plants collected from the wild population (plant ID 20.W). The two samples (A003 and A224) represent two different leaves of the same plant. Both bacteria were found from both samples and in no other sample. A003 had only small numbers of these two bacteria, 15 reads of *Candidatus* Sulcia muelleri and 24 reads of

Candidatus Nasuia deltocephalinicola. A224 had remarkably more: 393 reads of *Candidatus* Sulcia muelleri and 398 reads of *Candidatus* Nasuia deltocephalinicola. Easiest explanation for the finding is that a part of an insect ended up crushed with the leaf in the sample tube and hence DNA from the insect was also sequenced. However, no other signal from insect DNA was found. It also seems somewhat unlikely that samples from two different leaves of the same plant both ended up with insect DNA. Another speculative option is that these bacteria ended up inside the lead when an insect was feeding on the plant. This option, however, remains unconfirmed.

3.3 Community analysis

For the 65 less common bacterial genera, I performed community analysis to assess whether the plant genotype and the treatment in which it had been grown had any effect on the bacterial community within the leaf. I excluded the six most commonly found eubacterial genera (see above) from the analysis, because they were interpreted as being clearly not endophytic but originating from outside the plant, being either contaminants or pathogens.

I analysed the bacterial communities as presence-absence data only, i.e. the analysis only considers whether a given bacterial genera had been observed in a sample or not. The number of reads obtained was not considered. Furthermore, I pooled the samples together according to the treatment, resulting in a bacterial community data observed from a certain treatment. I could not perform this analysis from the mother plants, because after removing chloroplast, mitochondria and contaminant bacteria from the data, literally almost no observations of any bacteria from the mother plants were left. The community analysis therefore only considers the treatment in which the seedling was grown and the genotype (mother plant) and compares the effect it might have on the bacterial community of the seedlings.

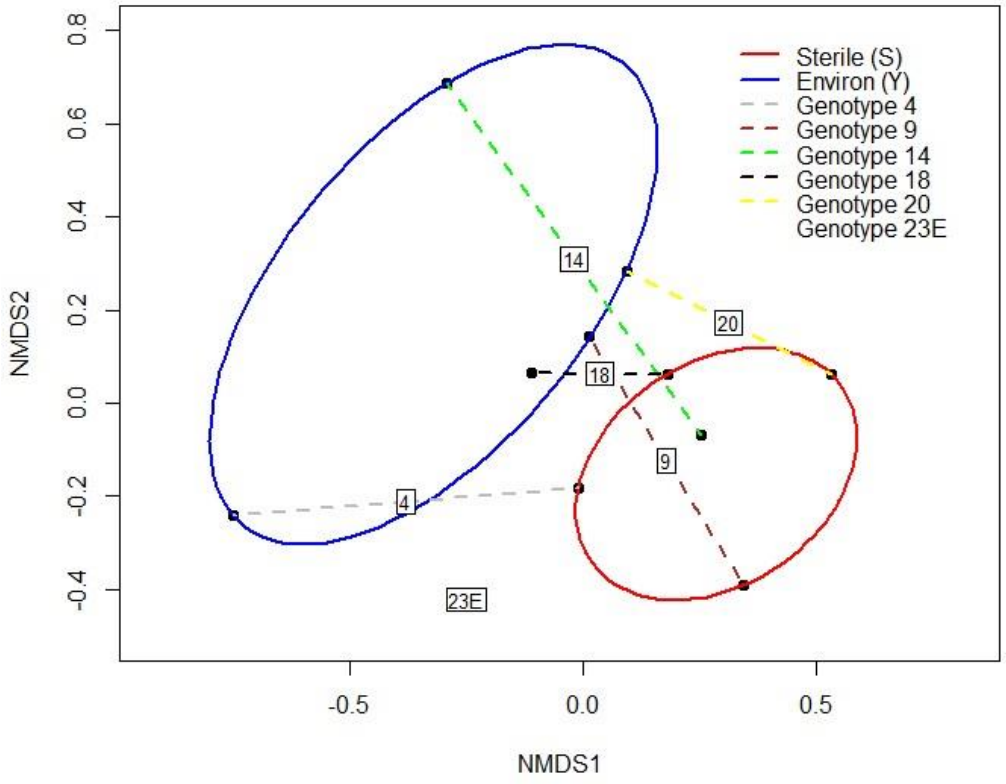
Bray-Curtis Dissimilarity values for all pairs of treatments are shown in Table 5. Bray-Curtis Dissimilarity index is a value between zero and one, zero meaning identical communities (low dissimilarity) and one meaning completely different communities (high dissimilarity) (Gardener 2014). Highest dissimilarity values (i.e. the most different bacterial communities) were calculated for treatment pairs 14.Y - 18.S (dissimilarity 0.782), 14.Y - 14.S (dissimilarity 0.750), 14.Y - 20.S (dissimilarity 0.740), 4.Y - 18.S (dissimilarity 0.737) and 4.S - 14.Y (dissimilarity 0.714). Notably, in all these cases the two groups are of different treatments, meaning a sterile group (S) versus an environmental group (Y) even if the genotype is the same. Lowest dissimilarity values (i.e. the most similar bacterial communities) were calculated for treatment pairs 14.S - 18.S (dissimilarity 0.270), 4.S - 14.S (dissimilarity 0.301), 18.Y - 20.Y (dissimilarity 0.301), 9.Y - 18.Y (dissimilarity

0.317) and 9.S - 20.S (dissimilarity 0.331). I found the lowest dissimilarity values between some of the genotypes belonging to the same treatment (sterile (S) versus sterile (S) or environmental (Y) versus environmental (Y)).

Table 5. Pairs of treatments and their corresponding Bray-Curtis dissimilarity values. *Five lowest values in* ***bold and italics****, five highest values in* **bold***. Number in treatment indicates the genotype, i.e. the ID number of the mother plant. Y = Environmental soil, S = Sterile soil, S2 = growth chamber control (e.g. 4.S = seeds from mother plant #4, grown in sterile soil.*

Pairs	Bray-Curtis	Pairs	Bray-Curtis	Pairs	Bray-Curtis
4.S - 4.Y	0.658	9.S - 9.Y	0.582	14.S - 20.Y	0.432
4.S - 9.S	0.410	9.S - 14.S	0.377	14.S - 23E.S2	0.619
4.S - 9.Y	0.457	9.S - 14.Y	0.709	**14.Y - 18.S**	**0.782**
4.S - 14.S	***0.301***	9.S - 18.S	0.455	14.Y - 18.Y	0.588
4.S - 14.Y	**0.714**	9.S - 18.Y	0.526	**14.Y - 20.S**	**0.740**
4.S - 18.S	0.357	***9.S - 20.S***	***0.331***	14.Y - 20.Y	0.600
4.S - 18.Y	0.368	9.S - 20.Y	0.523	14.Y - 23E.S2	0.583
4.S - 20.S	0.443	9.S - 23E.S2	0.586	18.S - 18.Y	0.458
4.S - 20.Y	0.440	9.Y - 14.S	0.495	18.S - 20.S	0.396
4.S - 23E.S2	0.576	9.Y - 14.Y	0.500	18.S - 20.Y	0.448
4.Y - 9.S	0.654	9.Y - 18.S	0.571	18.S - 23E.S2	0.649
4.Y - 9.Y	0.600	***9.Y - 18.Y***	***0.317***	18.Y - 20.S	0.538
4.Y - 14.S	0.702	9.Y - 20.S	0.600	***18.Y - 20.Y***	***0.301***
4.Y - 14.Y	0.593	9.Y - 20.Y	0.413	18.Y - 23E.S2	0.461
4.Y - 18.S	**0.737**	9.Y - 23E.S2	0.429	20.S - 20.Y	0.553
4.Y - 18.Y	0.529	**14.S - 14.Y**	**0.750**	20.S - 23E.S2	0.690
4.Y - 20.S	0.690	***14.S - 18.S***	***0.270***	20.Y - 23E.S2	0.508
4.Y - 20.Y	0.657	14.S - 18.Y	0.370		
4.Y - 23E.S2	0.610	14.S - 20.S	0.368		

I used the dissimilarity values described above to construct an ordination of the different treatments using Non-metric Multidimensional Scaling (NMDS). Ordination is a way of grouping similar communities together and graphically illustrating these groups. Picture 14 shows how bacterial communities of different treatments relate to each other in the ordination. The treatments that result closer to each other have relatively similar bacterial communities whereas groups that are further apart have more different bacterial communities. In the ordination plot the Sterile and Environmental groupings are clearly visible. I drew ellipses around the two soil treatment groups (Sterile (S) and Environmental (Y)) so that all points within that group are within the ellipse. In the resulting picture these two ellipses do not overlap, which highlights the fact that the bacterial communities of certain soil treatment tend to be closer to other communities within the same soil treatment groups. The growth chamber control group (23E) which had sterilized soil but was grown in the same growth chamber with the Environmental (Y) group, is not included in either ellipse but stands out separately in the ordination.



Picture 14. Ordination plot of different bacterial communities. *Ordination based on Bray-Curtis Dissimilarity values using Non-Metric Multidimensional Scaling (NMDS). Points represent bacterial communities in each treatment. Ordination plot shows graphically which communities are most similar according to Bray-Curtis Dissimilarity values. The greater the distance between two points, the more different the two communities are. Red ellipse contains all Sterile (S) treatments, blue ellipse contains all Environmental treatments (Y). The ellipses are simply drawn around the group so that all points in a given soil treatment group are within the corresponding ellipse. Dashed lines connect Sterile (S) and Environmental (Y) treatments of the same genotype (e.g. 4.S and 4.Y).*

I performed Analysis of Similarities (ANOSIM) to test the null hypothesis that there is no difference in the bacterial community composition between groups. I compared three sets of groups: 1) Soil Treatments (Sterile vs. Environmental), 2) Growth Chambers and 3) Genotypes (mother plant). ANOSIM results are given in Table 6. The ANOSIM statistic R is a value between -1 and 1. The closer it is to 1 the more similar the within group observations are to each other and the more dissimilar they are to observations from other groups.

Table 6. Analysis of similarities (ANOSIM) results for different groups.

Group	Significance (p)	ANOSIM statistic R
Soil treatment (S vs. Y)	0.014	0.3787
Growth chamber	0.011	0.5493
Genotype	0.964	-0.352

4 Discussion

In this study my aim was to answer the question about vertical inheritance of bacterial endophytes and how natural soil microbiome affects *Plantago lanceolata* germination, seedling growth and endophyte community. I first aimed to get a comprehensive picture of the overall endophyte community of *P. lanceolata* and then searched for clues of vertical inheritance, i.e. transmission of endophytes from mother to offspring via seed. I also studied the effect of the soil microbiome on germination, seedling growth and endophyte communities by performing a growth experiment with sterilized soil and environmental mixture soil. I included six *P. lanceolata* genotypes in my study to determine whether these patterns are genotype-specific.

4.1 Germination & seedling growth

I grew *P. lanceolata* seedlings in two different soil conditions: sterilized sand ("Sterile treatment" S) and sterilized sand mixed with soil collected from the field site ("Environmental treatment" Y). Soil treatment had a clear effect on the germination and growth of the seedlings (Picture 7 and 8). Seeds sown in sterile sand showed higher germination percentage than seeds of the same genotype grown in environmental soil, but these seedlings grew slower during the experiment and never reached the size that the seedlings grown in environmental mixture did. This effect was consistent across all genotypes. Genotype affected germination rate (Picture 7) but not the growth of the seedlings (Picture 8). For instance, the germination percentage of genotypes #18 and #20 was clearly higher in both the sterile and the environmental group than the germination percentage of genotypes #4, #9 and #14. Still within the genotype, germination percentage was consistently lower in the environmental soil treatment than in the sterile soil treatment across all genotypes (Picture 7).

Two important phenomena could explain the results of the growth experiment: 1) soil microbiota and 2) nutrients. In the sterile soil treatment, I tried to eliminate the natural soil microbiota as thoroughly as possible by autoclaving the sand twice. Even after this I could still see some fungal growth after I incubated the sand on a growth media for a few days. However, I observed no bacterial growth. Even though all of the microbiota could not be excluded from the "sterile" treatment, it is reasonable to expect remarkable reduction in the number of microbes compared to the soil collected from the field site since microbial biomass in natural soil is usually very high and autoclaving is an effective way to kill soil microbes (Carter, Yellowlees, Tibbett 2007). Natural microbiota (or its absence) could affect germination and seedling growth in at least two ways: 1) parasitic microbes (pathogens) could infect the emerging seedlings, since the plant is most

vulnerable to pathogens as a germinating seed when seed coat no longer protects the seed but its immune system is not yet fully operational (Agrios 1988). On the other hand, 2) beneficial microbes could strongly enhance seedling growth and survival (Ghimire, Charlton, Craven 2009; Wu et al. 2016). Furthermore, multicellular soil fauna could damage seeds and therefore deter growth (Brown and Gange 1989; Gange, Brown, Farmer 1991).

In my experiment I observed lower germination among seeds grown in environmental mixture soil but slower growth in sterile soil. Higher percentage of the seeds in sterile soil germinated but they also grew slower whereas a lower percentage of the seeds in environmental mixture germinated but they grew fast and reached greater sizes. It is possible that pathogenic microbes or other soil fauna reduced germination in environmental soil, while beneficial microbes enhanced growth among the seedlings that had managed to germinate. In the sterile soil, no pathogens would hinder germination but also no beneficial microbes would enhance seedling growth.

In this growth experiment it was difficult to control the nutrient conditions. At the start of the experiment, I applied no excess nutrients and did not measure the nutrient contents of the soil treatments. The environmental mixture soil probably had more available nutrients than the sterile soil. Five weeks after the start of the experiment I started applying additional fertilizing to both treatments, resulting in accelerated growth in both treatments. Hence it may be that part of the difference in growth speed of seedling was caused by difference in the initial nutrient conditions (McWilliam, Clements, Dowling 1970).

When it comes to the growth of test plants, growth chamber controls are clearly in line with the growth of the sterile soil group (Picture 8). That means that growth chamber had no effect on the growth of the seedlings. This suggests that the growth of the test plants during the experiment was mainly defined by the soil, most likely by its nutritional properties and maybe also by beneficial microbes, although neither of these was measured or characterized. The differences in the bacterial communities I have analysed below could have a link to the growth of the seedlings although no causal link can be inferred from this data. To further assess the effect of soil microbiota on the germination of seeds and survival of seedlings, it would be interesting to perform growth experiments with characterized soil microbiota or perhaps inoculating certain microbial strains into the soil. This approach has already been applied and it has been suggested as a useful tool in agriculture for biological disease-control and to promote healthy soil microbiota and beneficial plant-microbe interactions (Mendes, Garbeva, Raaijmakers 2013). Controlling nutritional conditions and pH of the soil would also be important in such an experiment.

4.2 OTU Analysis – difficulties in excluding mitochondrial DNA

Because of the prokaryotic origin of mitochondria and chloroplasts, there is a high risk of their genes interfering with studies looking at endophyte diversity inside plants. Mitochondria and chloroplasts also harbour 16S rDNA regions in their genomes, so the challenge is to exclude them from the analysis and only look at the eubacterial 16S rDNA regions. Furthermore, endophyte abundance is typically low in leaves (Hallmann et al. 1997). Leaves are also the most photosynthetically active parts of plants, hence having the highest number of chloroplasts. Combined, these two factors - low abundance of endophytes and presence of significant amounts of mitochondria and chloroplasts - make it extremely challenging to get comprehensive picture of the diversity of bacterial endophytes in leaves. No wonder the observations of bacterial endophytes were for a long time considered merely contaminants or latent pathogens (Holland and Polacco 1994).

The challenge of mitochondria and chloroplast DNA has been successfully overcome in previous endophyte studies by 1) designing fitting primers and by 2) size fractioning the obtained DNA products (Chelius and Triplett 2001; Kumar et al. 2017; Nissinen, Mannisto, van Elsas 2012). In this study, most Chloroplast sequences were successfully excluded with the appropriate primers. Some Chloroplast sequences still remained in the data, but this was expected and would not on its own have interfered dramatically with the analysis. Size fractioning of PCR products was however not done, despite the awareness of the risk of mitochondrial products overriding sequencing of endophyte DNA. Since no clearly separate endophyte and mitochondrial bands on gel electrophoresis were recognised, it was - apparently falsely - thought that the *P. lanceolata* mitochondrial 16S regions had not been amplified at all with the primers in use. This proved to be a major error of judgement, since 86.4% of all reads obtained ended up being mitochondrial sequences (Table 2, Picture 9). This highlights the importance of careful assessment of laboratory methods and results and possibly repeated preliminary analysis and pilot sequencing with a small number of samples before starting large scale sequencing. Simply redoing the analysis and carefully going through the size fractionation of initial PCR products could yield more reliable results.

Because the data in this study was dominated by mitochondrial and chloroplast sequences, extreme caution is needed when making inferences regarding the endophytic bacterial community in the leaves of *P. lanceolata*. The presence of mitochondria and chloroplast sequences effectively hinders the possibility to detect endophytic bacteria in these samples. This effect may even be exacerbated by the competitive nature of PCR and sequencing reactions. Each

consecutive PCR cycle will amplify the most common sequences in largest amounts. This results in the accumulation of most common sequences (in this case mitochondrial) and even lower relative abundance of less common sequences than in the original sample. Hence, if already in the original raw samples there was far more mitochondria and chloroplast sequences than bacterial sequences, the remaining bacterial sequences after PCR are even less representative of the total bacterial community. However, there should not be a particular bias between the genotypes nor the treatments in this effect, and hence below I will nevertheless describe the patterns of variation in bacterial communities detected in this study, while emphasizing the need for utmost caution when interpreting these results.

4.3 Eubacterial reads mainly from contaminants

After I removed mitochondria and chloroplast reads from the data, 38,481 DNA reads remained, representing only 1.1% of all reads obtained from sequencing (Table 2, Picture 9). Out of these 38,481 eubacterial DNA reads, 83% (31,981 reads) came from the six most common bacterial genera: *Shewanella*, *Methylobacterium*, *Ralstonia*, *Halomonas*, *Salinicola* and *Clostridium* (Table 3). The few eubacterial reads obtained, most notably the reads of the six most common genera, seem to originate most probably from contaminant bacteria rather than endophytes. Two of the six most common eubacterial genera, *Shewanella* and *Halomonas*, were found in all but one control samples: from empty extraction controls as well as empty sequencing controls (one sequencing control remaining free of any bacterial observations, Table 4). This seems to indicate that these signals come not from true plant associated bacteria but from contaminants. The source of contamination may be e.g. water or reagents used in laboratory processes.

Quite interestingly, in the case of one genera, *Ralstonia* sp., the contamination may have happened already in the DNA extraction phase: *Ralstonia* occurs in most samples that were extracted during the first 11 extraction days. Then suddenly from the 12th extraction day on, there remains no signal of *Ralstonia* except for very weak signals on the 13th and 15th day (Picture 12). This coincides with the opening of a new extraction kit: An old, previously opened kit had been used up until the 11th extraction day. On the 12th extraction day a new unopened kit was brought into use. Samples from the mother plants were processed during the first two extraction days. This may explain why *Ralstonia* occurrence is lower in the samples done during the first two days: leaves of mother plants probably hold higher amounts of chloroplasts and hence, for reasons described above, detection of bacterial strains is more difficult in these samples. Samples from the growth experiment were however not organised in any specific order relating to the genotype

or treatment during the DNA extraction, and hence it seems unlikely that the occurrence or *Ralstonia* could reflect genotypic differences or differences between treatments.

Occurrence of another common bacterial genus, *Salinicola*, coincides with the occurrence of the most common and most obvious contaminant, *Shewanella*. There is strong correlation between *Shewanella* occurrence and *Salinicola* occurrence (Picture 13). Furthermore, also *Salinicola* reads were observed from some empty control samples (table 4).

4.4 OTUs abundant in single samples: pathogens and insect symbionts?

Some bacterial OTUs occur only in a small number of samples, but in great amounts. *Methylobacterium* sp. is found in 60 samples but in the vast majority of these in minimal numbers (<70 reads/sample). However, in two samples high prevalence of *Methylobacterium* was observed (A059, 4,184 reads and A233, 3,537 reads). This might indicate infection by a single pathogen or contamination of the samples. *Methylobacterium* sp. are however not known as plant pathogens, but quite the opposite, as common phyllosphere symbionts (Knief et al. 2010). It is therefore possible that the observed *Methylobacterium* are actually a signal from the epiphytic microbes.

Clostridium sp. were found in only two samples: one of them has only two *Clostridium* reads but the other one 1065. Curiously, these two samples came from seedlings that were by accident sown in the same growth pot. Normally only one seed per growth pot was sown, but here two seeds ended up in the same pot. Furthermore, these seeds almost died during the experiment. They were noted as being “almost dead” by the sixth week of growth experiment but during the seventh week ended up growing new healthy leaves, from which the samples were taken. Thus, it remains possible that *Clostridium* bacteria had infected the plants, resulting in poor condition and growth of leaves.

Even two apparently insect-borne bacteria were found. Two samples contained two bacteria, *Candidatus* Sulcia muelleri and *Candidatus* Nasuia deltocephalinicola, that are known as obligate symbionts of leafhoppers (Bennett and Moran 2013). It is tempting to speculate about the possibility of insect symbionts gaining entry to plants as the insect feeds on the leaves. This happens frequently when viral pathogens gain access to plants by insect vectors (Agrios 1988). Another explanation in this case, however, could be that some remnants of an insect or perhaps inset eggs have ended up in the sample and being sequenced alongside the plant material.

4.5 Bacterial communities

I decided to perform the community analysis without the six most commonly observed bacterial genera. The rationale for this is that these six most common bacterial genera were most probably non-endophytic. This can be said with good confidence for reasons described above. Furthermore, had these most likely non-endophytic bacteria been included in the analysis, they would have overruled the analysis so that the bacterial communities (and the degree of their similarity) would have been mostly defined by these most commonly observed genera.

After removing these six genera and a few unidentified reads, 166 different OTUs remained, representing 65 different bacterial genera. Community analysis using Bray-Curtis Dissimilarity values and ordination analysis seems to indicate that plant genotype had no observable effect on the bacterial community. Dissimilarity values for the bacterial communities did not associate with the plant genotype. Communities of the same genotype got sometimes high, sometimes low dissimilarity values. In the ordination, plants of the same genotype do not cluster together. It therefore seems, that in this study, genotype does not play a major role in defining bacterial community. Species-specific differences in endophyte assembly have been observed previously (Kumar et al. 2017), so perhaps intraspecific genotypic differences in the endophyte assembly are too small to detect here. It would be interesting to analyse groups of closely related plant species and their endophyte community to shed light on the patterns of coevolution.

Unlike genotypes, the soil treatments seem to cluster the bacterial communities together: In all five treatment pairs with lowest dissimilarity values (i.e. the most similar bacterial communities), the two members of the pair were of the same soil treatment. Analysis of Similarities (ANOSIM) also gave significant results ($p = 0.014$) for the test according to the soil treatment. ANOSIM statistic R, is however only 0.3787, which indicates that the effect is not very strong. This is understandable by looking at the same analysis with samples grouped according to the growth chamber. Here ANOSIM also yields significant results ($p = 0.011$) and the test statistic R is higher (0.5493), meaning that the *growth chamber* seems to have bigger effect on the similarities of bacterial communities than the soil treatment. Soil treatments were kept in separate growth chambers to minimize the risk of contamination from environmental soil into the sterilized soil. Some seeds were grown in sterilized soil but in the same growth chamber as the environmental soil treatment, to act as growth chamber controls. Now it does actually seem that when the growth chamber controls are included in the analysis as part of the sterile treatment group, the bacterial communities group better together according to the growth chamber than according to the soil treatment. Perhaps contamination did in fact happen during the growth experiment,

resulting in microbes from the environmental soil mixture to affect the bacterial community of growth chamber control plants. In the ordination plot (Picture 14) the growth chamber control group (23E) stands somewhat separate from both Sterile group and Environmental group, but there are also remarkable distances within both Sterile and Environmental group.

Although some differences could be observed in the community analysis, I want to emphasize that these results need to be considered with caution. Because of the difficulties in excluding mitochondrial DNA from the data, my view of the bacterial communities is most likely inaccurate. To get better results, the first thing to do would be to carefully develop the method I have used to get a reliable view of the bacterial communities inside the leaves of *P. lanceolata*. If the number of reads from mitochondrial DNA could be reduced remarkably, the data could turn out to be more useful.

5 Conclusions

I have analysed here the leaf endophytic bacterial community of *P. lanceolata* and factors affecting its composition including maternal genotype and soil conditions. In addition, I have analysed how soil properties affect the germination and growth of *P. lanceolata* seedlings.

I was not able to answer reliably to the main questions of this study, because of methodological challenges. However, useful lessons have been learned that could point towards new research directions. Validating the molecular methods for studying *P. lanceolata* endophytes could open up interesting possibilities for studying coevolution of host-endophyte interactions and the ways in which environmental conditions affect endophyte communities. Repeating DNA sequencing with the primers used in this study and going through careful size fractionation and nested PCR on the acquired DNA products could yield more reliable data. If a reliable snapshot view of the endophytic bacteria could be acquired, the next step could involve more elaborate experimental setups, somewhat like the growth experiment I have described in this study. However, starting time consuming and resource-intensive experiments and analysis are better left until after the methods are duly validated.

In this study I observed clear differences in the germination and growth of *P. lanceolata* seedlings in different soil treatments as well as their subsequent growth during the seven weeks after germination. These differences are probably due to soil microbiota and other soil fauna and nutritional conditions. I also found out that many samples were contaminated during the laboratory analysis by bacteria that were identified as being clearly non-endophytic. After removing such contaminants from the analysis, I was able to detect some differences in the bacterial communities observed in different samples, some potentially attributable to the different soil treatments. These results, however, can only be considered preliminary and would require further validation.

Acknowledgements

I would like to thank my supervisors Anna-Liisa Laine, Riitta Nissinen and Hanna Susi for their support and encouragement. They have taught me a lot during all phases of the project from the initial ideas, through experimental design, field work, analysis and writing. I thank Hanna Susi and Suvi Sallinen for their help and company during the field work and Hanna Susi and Laura Häkkinen for their help and guidance during laboratory work. Krista Raveala helped a lot during many phases of the practical work. My sincerest gratitude to her. I acknowledge the European Research Council for funding (Consolidator Grant 'RESISTANCE' to Anna-Liisa Laine).

References

Afzal I, Shinwari ZK, Sikandar S, Shahzad S. 2019. Plant beneficial endophytic bacteria: Mechanisms, diversity, host range and genetic determinants. Microbiological Research. 221:36-49.

Agrios GN. 1988. Introduction to plant pathology. Plant Pathology, 3rd Ed.Academic Press, San Diego, CA. :3-40.

Araújo WL, Marcon J, Maccheroni WJ, van Elsas JD, van Vuurde, Jim W L, Azevedo JL. 2002. Diversity of endophytic bacterial populations and their interaction with xylella fastidiosa in citrus plants. Applied and Environmental Microbiology. 68(10):4906-14.

Baker GC, Smith JJ, Cowan DA. 2003. Review and re-analysis of domain-specific 16S primers. J Microbiol Methods. 55(3):541-55.

Barret M, Briand M, Bonneau S, Préveaux A, Valière S, Bouchez O, Hunault G, Simoneau P, Jacquesa M. 2015. Emergence shapes the structure of the seed microbiota. Applied and Environmental Microbiology. 81(4):1257-66.

Bashan Y, Holguin G. 1998. Proposal for the division of plant growth-promoting rhizobacteria into two classifications: Biocontrol-PGPB (plant growth-promoting bacteria) and PGPB. Soil Biology and Biochemistry. 30(8):1225-8.

Bennett GM, Moran NA. 2013. Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect. Genome Biology and Evolution. 5(9):1675-88.

Berendsen RL, Pieterse CMJ, Bakker PAHM. 2012. The rhizosphere microbiome and plant health. Trends in Plant Science. 17(8):478.

Berg G, Rybakova D, Grube M, Köberl M. 2016. The plant microbiome explored: Implications for experimental botany. Journal of Experimental Botany. 67(4):995-1002.

Berlec A. 2012. Novel techniques and findings in the study of plant microbiota: Search for plant probiotics. Plant Science : An International Journal of Experimental Plant Biology. 193-194:96-102.

Bhattacharyya P, Jha D. 2012. Plant growth-promoting rhizobacteria (PGPR): Emergence in agriculture. World J Microbiol Biotechnol. 28(4):1327-50.

Brown VK, Gange AC. 1989. Herbivory by soil-dwelling insects depresses plant species richness. Functional Ecology. 3(6):667-71.

Cankar K, Kraigher H, Ravnikar M, Rupnik M. 2005. Bacterial endophytes from seeds of norway spruce (picea abies L. karst). FEMS Microbiology Letters. 244(2):341-5.

Carlier A, Fehr L, Pinto-Carbó M, Schäberle T, Reher R, Dessein S, König G, Eberl L. 2016. The genome analysis of Candidatus Burkholderia crenata reveals that secondary metabolism may be a key function of the ardisia crenata leaf nodule symbiosis. Environmental Microbiology. 18(8):2507-22.

Carter DO, Yellowlees D, Tibbett M. 2007. Autoclaving kills soil microbes yet soil enzymes remain active. Pedobiologia - International Journal of Soil Biology. 51(4):295-9.

Chelius MK, Triplett EW. 2001. The diversity of archaea and bacteria in association with the roots of zea mays L. Microb Ecol. 41(3):252-63.

Chimwamurombe PM, Grönemeyer JL, Reinhold-Hurek B, Mitter B. 2016. Isolation and characterization of culturable seed-associated bacterial endophytes from gnotobiotically grown marama bean seedlings. FEMS Microbiology Ecology. 92(6):fiw083.

Cho I, Blaser MJ. 2012. The human microbiome: At the interface of health and disease. Nature Reviews. Genetics. 13(4):260-70.

Compant S, Clément C, Sessitsch A. 2010. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. Soil Biology and Biochemistry. 42(5):669-78.

Compant S, Saikkonen K, Mitter B, Campisano A, Mercado-Blanco J. 2016. Editorial special issue: Soil, plants and endophytes. .

Cope-Selby N, Cookson A, Squance M, Donnison I, Flavell R, Farrar K. 2017. Endophytic bacteria in miscanthus seed: Implications for germination, vertical inheritance of endophytes, plant evolution and breeding. GCB Bioenergy. 9(1):57-77.

Ding T, Melcher U. 2016. Influences of plant species, season and location on leaf endophytic bacterial communities of non-cultivated plants. PloS One. 11(3):e0150895.

Ferreira A, Quecine MC, Lacava PT, Oda S, Azevedo JL, Araújo WL. 2008. Diversity of endophytic bacteria from eucalyptus species seeds and colonization of seedlings by pantoea agglomerans. FEMS Microbiology Letters. 287(1):8.

Franche C, Lindström K, Elmerich C. 2009. Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. Plant and Soil. 321(1/2):35-59.

Frank AC, Guzmán JPS, Shay JE. 2017. Transmission of bacterial endophytes. Microorganisms. 5(4):70.

Gange AC, Brown VK, Farmer LM. 1991. Mechanisms of seedling mortality by subterranean insect herbivores. Oecologia. 88(2):228-32.

Gardener M. 2014. Community ecology: Analytical methods using R and excel. Exeter [United Kingdom]: Pelagic Publishing.

Gehrig H, Schüssler A, Kluge M. 1996. Geosiphon pyriforme, a fungus forming endocytobiosis with nostoc (cyanobacteria), is an ancestral member of the glomales: Evidence by SSU rRNA analysis. Journal of Molecular Evolution. 43(1):71.

Ghimire S, Charlton N, Craven K. 2009. The mycorrhizal fungus, sebacina vermifera, enhances seed germination and biomass production in switchgrass (panicum virgatum L). Bioenerg Res. 2(1):51-8.

Ghyselinck J, Pfeiffer S, Heylen K, Sessitsch A, De Vos P. 2013. The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies. PLoS One. 8(8):e71360.

Hallmann J, Quadt-Hallmann A, Mahaffee WF, Kloepper JW. 1997. Bacterial endophytes in agricultural crops. Canadian Journal of Microbiology. 43(10):895-914.

Hanski IA. 2011. Eco-evolutionary spatial dynamics in the glanville fritillary butterfly. Proceedings of the National Academy of Sciences of the United States of America. 108(35):14397-404.

Hardoim PR, Hardoim CCP, van Overbeek LS, van Elsas JD. 2012. Dynamics of seed-borne rice endophytes on early plant growth stages. PLoS ONE. 7(2):e30438.

Hardoim PR, van Overbeek LS, van Elsas JD. 2008. Properties of bacterial endophytes and their proposed role in plant growth. Trends in Microbiology. 16(10):463-71.

Hardoim PR, van Overbeek LS, Berg G, Pirttilä AM, Compant S, Campisano A, Döring M, Sessitsch A. 2015. The hidden world within plants: Ecological and evolutionary considerations for defining functioning of microbial endophytes. Microbiology and Molecular Biology Reviews. 79(3):293-320.

Head IM, Saunders JR, Pickup RW. 1998. Microbial evolution, diversity, and ecology: A decade of ribosomal RNA analysis of uncultivated microorganisms. Microb Ecol. 35(1):1-21.

Heijden MGA, Martin FM, Selosse M, Sanders IR. 2015. Mycorrhizal ecology and evolution: The past, the present, and the future. New Phytologist. 205(4):1406-23.

Hironobu M, Morisaki H. 2007. Endophytic bacteria in the rice plant. The FEBS Journal. 274(20):5223.

Hodgson S, Cates C, Hodgson J, Morley NJ, Sutton BC, Gange AC. 2014. Vertical transmission of fungal endophytes is widespread in forbs. Ecology and Evolution. 4(8):1199-208.

Holland MA, Polacco JC. 1994. PPFMs and other covert contaminants: Is there more to plant physiology than just plant? Annual Review of Plant Physiology and Plant Molecular Biology. 45(1):197-209.

Honegger R. 1991. Functional aspects of the lichen symbiosis. Annual Review of Plant Physiology and Plant Molecular Biology. 42(1):553-78.

Izumi H, Anderson IC, Killham K, Moore ERB. 2008. Diversity of predominant endophytic bacteria in european deciduous and coniferous trees. Canadian Journal of Microbiology. 54(3):173-9.

Jones P, Garcia BJ, Furches A, Tuskan GA, Jacobson D. 2019. Plant host-associated mechanisms for microbial selection. Frontiers in Plant Science. 10(N/A):862.

Jousimo J, Tack AJM, Ovaskainen O, Mononen T, Susi H, Tollenaere C, Laine A. 2014. Disease ecology. ecological and evolutionary effects of fragmentation on infectious disease dynamics. Science (New York, N.Y.). 344(6189):1289.

Knief C, Ramette A, Frances L, Alonso-Blanco C, Vorholt JA. 2010. Site and plant species are important determinants of the methylobacterium community composition in the plant phyllosphere. The ISME Journal. 4(6):719-28.

Kumar M, Brader G, Sessitsch A, Mäki A, van Elsas JD, Nissinen R. 2017. Plants assemble species specific bacterial communities from common core taxa in three arcto-alpine climate zones. Frontiers in Microbiology. 8.

Laine A, Hanski I. 2006. Large-scale spatial dynamics of a specialist plant pathogen in a fragmented landscape. Journal of Ecology. 94(1):217-26.

Laitinen RK, Hellström KO, Wäli PR. 2016. Context-dependent outcomes of subarctic grass-endophyte symbiosis. Fungal Ecology. 23:66-74.

Leigh EG. 2010. The evolution of mutualism. Journal of Evolutionary Biology. 23(12):2507-28.

Lemaire B, Lachenaud O, Persson C, Smets E, Dessein S. 2012. Screening for leaf-associated endophytes in the genus psychotria (rubiaceae). FEMS Microbiology Ecology. 81(2):364-72.

Lemaire B, Vandamme P, Merckx V, Smets E, Dessein S. 2011. Bacterial leaf symbiosis in angiosperms: Host specificity without co-speciation. PloS One. 6(9):e24430.

Li J, Zhao G, Varma A, Qin S, Xiong Z, Huang H, Zhu W, Zhao L, Xu L, Zhang S, et al. 2012. An endophytic pseudonocardia species induces the production of artemisinin in artemisia annua. PloS One. 7(12):e51410.

Lopez B, Bashan Y, Bacilio M. 2011. Endophytic bacteria of mammillaria fraileana, an endemic rock-colonizing cactus of the southern sonoran desert. Arch Microbiol. 193(7):527-41.

López MA, Bannenberg G, Castresana C. 2008. Controlling hormone signaling is a plant and pathogen challenge for growth and survival. Current Opinion in Plant Biology. 11(4):420-7.

Lunau K. 2004. Adaptive radiation and coevolution — pollination biology case studies. Organisms Diversity & Evolution. 4(3):207-24.

Marquez LM, Redman RS, Rodriguez RJ, Roossinck MJ. 2007. A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. Science. 315(5811):513-5.

Masson-Boivin C, Giraud E, Perret X, Batut J. 2009. Establishing nitrogen-fixing symbiosis with legumes: How many rhizobium recipes? Trends in Microbiology. 17(10):458-66.

McCutcheon JP, Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 my of evolution. Genome Biology and Evolution. 2:708-18.

McWilliam JR, Clements RJ, Dowling PM. 1970. Some factors influencing the germination and early seedling development of pasture plants. Crop & Pasture Science. 21(1):19-32.

Mendes R, Garbeva P, Raaijmakers JM. 2013. The rhizosphere microbiome: Significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. FEMS Microbiol Rev. 37(5):634-63.

Miliute I, Buzaite O, Baniulis D, Stanys V. 2015. Bacterial endophytes in agricultural crops and their role in stress tolerance: A review. Zemdirbyste-Agriculture. 102(4):465-78.

Moran NA, McLaughlin HJ, Sorek R. 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science. 323(5912):379-82.

Morris JJ, Lenski RE, Zinser ER. 2012. The black queen hypothesis: Evolution of dependencies through adaptive gene loss. mBio. 3(2).

Nissinen RM, Mannisto MK, van Elsas JD. 2012. Endophytic bacterial communities in three arctic plants from low arctic fell tundra are cold-adapted and host-plant specific. FEMS Microbial Ecology. 82(2):510-22.

Noda H, Watanabe K, Kawai S, Yukuhiro F, Miyoshi T, Tomizawa M, Koizumi Y, Nikoh N, Fukatsu T. 2012. Bacteriome-associated endosymbionts of the green rice leafhopper nephotettix cincticeps (hemiptera: Cicadellidae). Appl Entomol Zool. 47(3):217-25.

Ojanen SP, Nieminen M, Meyke E, Pöyry J, Hanski I. 2013. Long-term metapopulation study of the glanville fritillary butterfly (melitaea cinxia): Survey methods, data management, and long-term population trends. Ecology and Evolution. 3(11):3713-37.

Ortíz-Castro R, Contreras-Cornejo HA, Macías-Rodríguez L, López-Bucio J. 2009. The role of microbial signals in plant growth and development. Plant Signaling & Behavior. 4(8):701-12.

Peñuelas J, Terradas J. 2014. The foliar microbiome. Trends in Plant Science. 19(5):278-80.

Philippot L, Raaijmakers JM, Lemanceau P, Van der Putten WH. 2013. Going back to the roots: The microbial ecology of the rhizosphere. Nature Reviews Microbiology. 11(11):789-99.

Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Bakker PAHM. 2014. Induced systemic resistance by beneficial microbes. Annual Review of Phytopathology. 52(1):347-75.

Pinto-Carbó M, Sieber S, Dessein S, Wicker T, Verstraete B, Gademann K, Eberl L, Carlier A. 2016. Evidence of horizontal gene transfer between obligate leaf nodule symbionts. The ISME Journal. 10(9):2092-105.

Pirozynski KA, Malloch DW. 1975. The origin of land plants: A matter of mycotrophism. Biosystems. 6(3):153-64.

Puente EM, Li CY, Bashan Y. 2009. Rock-degrading endophytic bacteria in cacti. Environmental and Experimental Botany. 66(3):389-401.

Raaijmakers JM, Paulitz TC, Steinberg C, Alabouvette C, Moënne-Loccoz Y. 2009. The rhizosphere: A playground and battlefield for soilborne pathogens and beneficial microorganisms. Plant and Soil. 321:1-2.

Rosenberg E, Zilber-Rosenberg I. 2016. Microbes drive evolution of animals and plants: The hologenome concept. mBio. 7(2):e01395.

Rosenberg E, Sharon G, Atad I, Zilber-Rosenberg I. 2010. The evolution of animals and plants via symbiosis with microorganisms. Environmental Microbiology Reports. 2(4):500-6.

Rosenblueth M, Martínez-Romero E. 2006. Bacterial endophytes and their interactions with hosts. Molecular Plant-Microbe Interactions : MPMI. 19(8):827-37.

Rubin R, van Groenigen K, Hungate B. 2017. Plant growth promoting rhizobacteria are more effective under drought: A meta-analysis. Plant Soil. 416(1):309-23.

Sachs JL, Skophammer RG, Regus JU. 2011. Evolutionary transitions in bacterial symbiosis. Proceedings of the National Academy of Sciences of the United States of America. 108(Supplement 2):10800-7.

Sánchez-Cañizares C, Jorrín B, Poole PS, Tkacz A. 2017. Understanding the holobiont: The interdependence of plants and their microbiome. Current Opinion in Microbiology. 38:188-96.

Santoyo G, Moreno-Hagelsieb G, Orozco-Mosqueda MdC, Glick BR. 2016. Plant growth-promoting bacterial endophytes. Microbiological Research. 183:92-9.

Selosse M, Le Tacon F. 1998. The land flora: A phototroph-fungus partnership? Trends in Ecology & Evolution. 13(1):15-20.

Selosse M, Bessis A, Pozo MJ. 2014. Microbial priming of plant and animal immunity: Symbionts as developmental signals. Trends in Microbiology. 22(11):607-13.

Sieber S, Carlier A, Neuburger M, Grabenweger G, Eberl L, Gademann K. 2015. Isolation and total synthesis of kirkamide, an aminocyclitol from an obligate leaf nodule symbiont. Angewandte Chemie International Edition. 54(27):7968-70.

Spaepen S, Vanderleyden J. 2011. Auxin and plant-microbe interactions. Cold Spring Harbor Perspectives in Biology. 3(4):a001438.

Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat, Emiel Ver Loren, Brown JKM, Butcher SA, Gurr SJ, et al. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science. 330(6010):1543-6.

Stoye J, P. 2012. Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nature Reviews. Microbiology. 10(6):395-406.

Strange RN, Scott PR. 2005. Plant disease: A threat to global food security. Annual Review of Phytopathology. 43(1):83-116.

Sturz AV, Christie BR, Matheson BG, Nowak J. 1997. Biodiversity of endophytic bacteria which colonize red clover nodules, roots, stems and foliage and their influence on host growth. Biol Fertil Soils. 25(1):13-9.

Tello S, Silva-Flores P, Agerer R, Halbwachs H, Beck A, Peršoh D. 2014. Hygrocybe virginea is a systemic endophyte of plantago lanceolata. Mycol Progress. 13(3):471-5.

Truyens S, Weyens N, Cuypers A, Vangronsveld J. 2015. Bacterial seed endophytes: Genera, vertical transmission and interaction with plants. Environmental Microbiology Reports. 7(1):40-50.

Turner TR, James EK, Poole PS. 2013. The plant microbiome. Genome Biology. 14(6):209.

Van Elst D, Nuyens S, van Wyk B, Verstraete B, Dessein S, Prinsen E. 2013. Distribution of the cardiotoxin pavettamine in the coffee family (rubiaceae) and its significance for gousiekte, a fatal poisoning of ruminants. Plant Physiology and Biochemistry. 67:15-9.

Vega FE, Pava-Ripoll M, Posada F, Buyer JS. 2005. Endophytic bacteria in coffea arabica L. Journal of Basic Microbiology. 45(5):371.

Verstraete B, Van Elst D, Steyn H, Van Wyk B, Lemaire B, Smets E, Dessein S. 2011. Endophytic bacteria in toxic south african plants: Identification, phylogeny and possible involvement in gousiekte. PLoS One. 6(4):e19265.

Wu Y, Feng Y, Paré PW, Chen Y, Xu R, Wu S, Wang S, Zhao Q, Li H, Wang Y, et al. 2016. Beneficial soil microbe promotes seed germination, plant growth and photosynthesis in herbal crop codonopsis pilosula. Crop and Pasture Science. 67(1):91.

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nature Reviews Microbiology. 12(9):635.

Zheng D, Alm EW, Stahl DA, Raskin L. 1996. Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. Applied and Environmental Microbiology. 62(12):4504-13.

Zilber-Rosenberg I, Rosenberg E. 2008. Role of microorganisms in the evolution of animals and plants: The hologenome theory of evolution. FEMS Microbiology Reviews. 32(5):723-35.